

### Analysis Scripts in Large-Scale Assessments in Education: Scripts based on PIAAC data

Veröffentlichungsversion / Published Version  
Sammelwerk / collection

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:  
GESIS - Leibniz-Institut für Sozialwissenschaften

**Empfohlene Zitierung / Suggested Citation:**

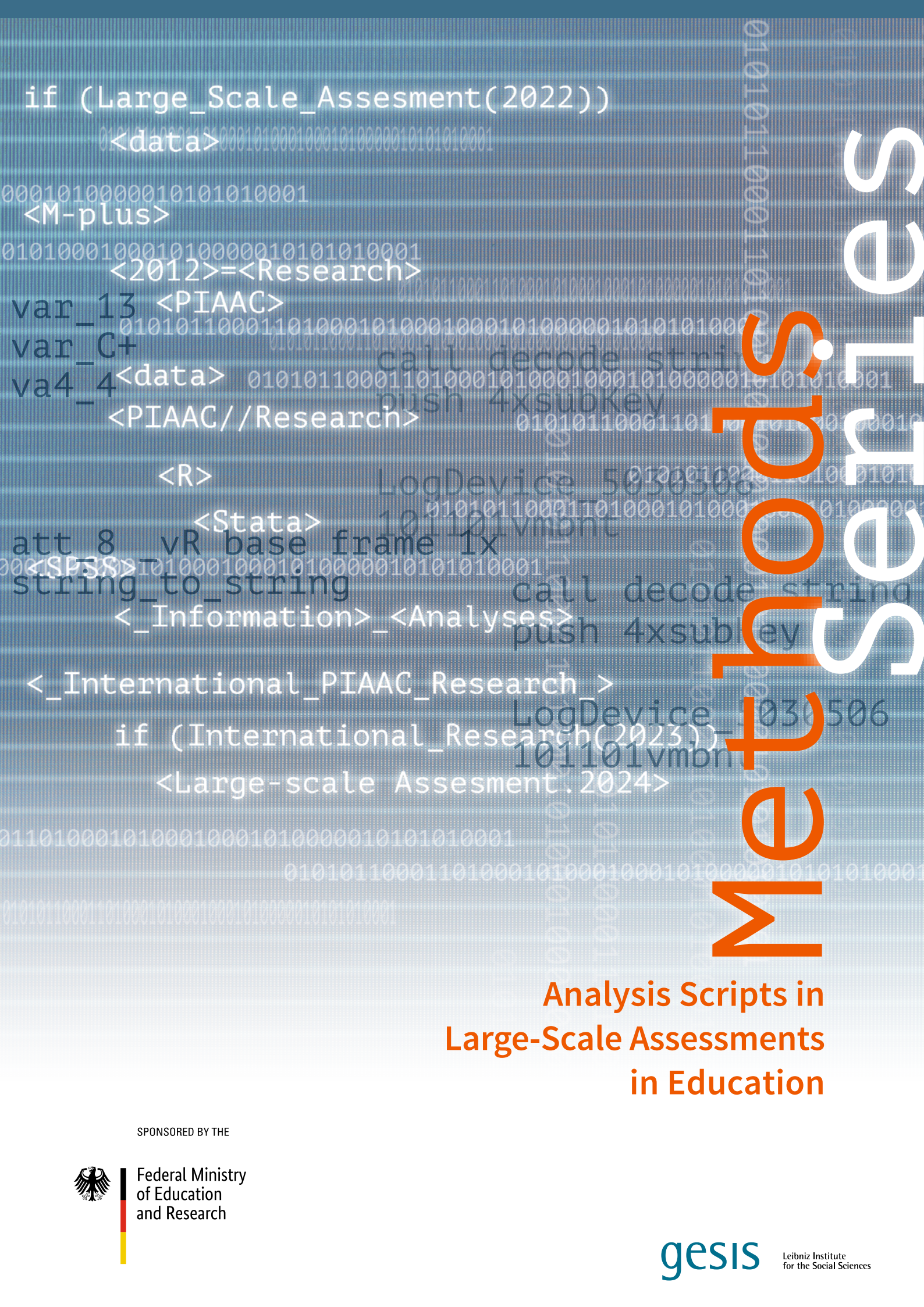
Research Data Center PIAAC at GESIS. (2023). *Analysis Scripts in Large-Scale Assessments in Education: Scripts based on PIAAC data*. (Methods Series). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/ssoar.85846>

**Nutzungsbedingungen:**

Dieser Text wird unter einer CC BY-NC Lizenz (Namensnennung-Nicht-kommerziell) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by-nc/4.0/deed.de>

**Terms of use:**

This document is made available under a CC BY-NC Licence (Attribution-NonCommercial). For more information see: <https://creativecommons.org/licenses/by-nc/4.0>



```
if (Large_Scale_Assesment(2022))
```

```
  <data>
```

```
  <M-plus>
```

```
  <2012>=<Research>
```

```
var_13 <PIAAC>
```

```
var_C+
```

```
va4_4 <data>
```

```
  <PIAAC//Research>
```

```
  <R>
```

```
  <Stata>
```

```
att_8_vR base frame ix
```

```
<SPSS> string_to_string
```

```
  <_Information>_<Analyses>
```

```
<_International_PIAAC_Research >
```

```
  if (International_Research(2023))
```

```
    <Large-scale Assesment.2024>
```

# SCRIPTS FOR ASSESSMENT

## Analysis Scripts in Large-Scale Assessments in Education

SPONSORED BY THE



Federal Ministry of Education and Research

gesis

Leibniz Institute for the Social Sciences

# Methods Series

## Analysis Scripts in Large-Scale Assessments in Education

**Research Data Center PIAAC  
(RDC PIAAC) at GESIS**

**Contact:** [fdz-piaac@gesis.org](mailto:fdz-piaac@gesis.org)  
**Webpage:** <https://www.gesis.org/en/piaac/rdc>

## Scientific Advisory Board

---

The Scientific Advisory Board of the Methods Series will review scripts submitted for inclusion in future volumes of the series. Made up of researchers and experts in the field with a methodological focus and/or expertise in PIAAC data analyses, the board currently comprises the following researchers (alphabetical order):

Paul Bailey  
(American Institutes for Research, Arlington, VA, United States)

Julia Gorges  
(University of Marburg, Germany)

Jan Paul Heisig  
(Berlin Social Science Center [WZB], Germany)

Ronny Scherer  
(Centre for Educational Measurement at the University of Oslo, Norway)

Rolf Strietholt  
(International Association for the Evaluation of Educational Achievement IEA, Germany)

Simon Wiederhold  
(Catholic University Eichstaett-Ingolstadt, Germany)

Ting Zhang  
(American Institutes for Research, Arlington, VA, United States)

## Content

<b>Scientific Advisory Board</b>	<b>3</b>
<b>1. PIAAC Methods Series: An Introduction</b>	<b>6</b>
1.1 PIAAC Analysis Scripts	6
1.2 Availability of PIAAC Data	7
1.3 PIAAC Methodology: Analyzing PIAAC Data	11
1.4 PIAAC Discovery: A Shiny App for Visualizing PIAAC Data	15
1.5 References	16
<b>2 Analyzing PIAAC data with Stata</b>	<b>18</b>
2.1 Introduction	18
2.2 Data Access and First Steps in Stata	18
2.3 Analyzing PIAAC Data Using the Package PIAACTOOLS	20
2.3.1 Examples <i>piaacdes</i>	20
2.3.2 Examples <i>piaactab</i>	21
2.3.3 Examples <i>piaacreg</i>	21
2.4 Analyzing PIAAC Data Using the REPEST Macro	22
2.4.1 Examples Descriptive Statistics	22
2.4.2 Examples Regression Analysis	24
2.5 References	24
<b>3 Analyzing PIAAC Data With R</b>	<b>28</b>
3.1 Introduction	28
3.2 Data Preparation and First Steps in R	28
3.2.1 Loading and Reading in Data	28
3.2.2 Data Exploration Using EdSurvey	29
3.3 Descriptive Statistics	29
3.3.1 Correlation	30
3.4 Data Manipulation	30
3.5 Data Analysis	31
3.5.1 Linear Regression	31
3.5.2 Logistic Regression	31
3.5.3 Gap Analysis	31
3.5.4 Percentile Analysis	32
3.5.5 Proficiency Level Analysis	32

3.6	Notes.....	32
3.7	References.....	33
<b>4</b>	<b>Analyzing PIAAC Data With Mplus.....</b>	<b>34</b>
4.1	Introduction.....	34
4.2	Structural Models With PIAAC Data.....	34
4.2.1	Path Model CURIOUS-LIFE-PSTRE.....	34
4.2.2	Multi-Group Path Model CURIOUS-LIFE-PSTRE.....	36
4.2.3	Multi-Group Path Model CURIOUS-LIFE-PSTRE With Equal <b>a</b> - and <b>b</b> -Paths Across Gender.....	37
4.2.4	Multi-Group Path Model CURIOUS-LIFE-PSTRE With Equal Indirect Effect Across Gender.....	38
4.3	Structural Equation Models With PIAAC Data.....	39
4.3.1	SEM With Measurement Models and Complex Structural Relations.....	39
4.3.2	Multi-Group SEM With Measurement Models, Complex Structural Relations, and Configural Measurement and Regression Invariance.....	42
4.3.3	Multi-Group SEM With Measurement Models, Complex Structural Relations, and Metric Measurement and Regression Invariance.....	46
4.3.4	Multi-Group SEM With Measurement Models, Complex Structural Relations, and Metric Measurement and Configural Regression Invariance.....	48
4.4	References.....	51
<b>5</b>	<b>Using the IDB Analyzer to Analyze PIAAC Data With SPSS, SAS, and R.....</b>	<b>52</b>
5.1	Introduction.....	52
5.2	Computing Means With Plausible Values.....	52
5.2.1	SPSS.....	53
5.2.2	SAS.....	53
5.2.3	R.....	55
5.3	Computing Linear Regression With Plausible Values As An Explanatory Variable.....	56
5.3.1	SPSS.....	56
5.3.2	SAS.....	57
5.3.3	R.....	59
5.4	References.....	60

# 1. PIAAC Methods Series: An Introduction

*Débora B. Maehler, Robin-Kiara Braun, Daria Morozova, Jennifer Dickson & Dennis Schüle  
(GESIS – Leibniz-Institute for the Social Sciences, Germany)*

## 1.1 PIAAC Analysis Scripts

The Programme for the International Assessment of Adult Competencies (PIAAC) is an international large-scale study initiated by the Organisation for Economic Co-operation and Development (OECD). PIAAC assesses key cognitive skills of the adult population in more than 40 countries. The skills assessed are literacy (the ability to understand, use, and interpret written texts; Jones et al., 2009), numeracy (the ability to retrieve, interpret, and use mathematical information; Gal et al., 2009), and problem solving in technology-rich environments (PSTRE; the ability to use technology to access and process information; Rouet et al., 2009). These skills are considered essential to successfully navigate demands in everyday life and in the workplace (OECD, 2013; OECD, 2021). Besides assessing cognitive skills, PIAAC also collects extensive background and job-related information, for example, gender; age; language most often spoken at home; number of people living in the household; number of books at home; parental information; children; education (highest school qualification, highest professional qualification, continuing education and training); work (employment status, occupation and industry, information on current job, information on last job, job search, years of paid work during lifetime, earnings); learning strategies; political efficacy; social trust; cultural engagement; skills mismatch; and skill use (e.g., literacy, numeracy, computer skills). An overview of the extent to which the data have been used in research to date can be found in the PIAAC bibliography (Maehler & Konradt, 2022).

Compared with other social science data, the large-scale PIAAC data are very complex, which makes them difficult to analyze without prior knowledge. For example, they contain multiple competence estimates per person (10 plausible values) and a different number of weighting factors per country, which must be taken into account in the analyses. In this context, applying the appropriate analytical method and the available calculation procedures is a challenge, especially for early career researchers, as it is time-consuming and can be a significant source of error in published results. Furthermore, the growing interdisciplinarity of research has changed the research landscape: The methods used are not solely discipline-specific, and the accessibility of the software and the analysis spectrum (i.e., the extent to which a question can be processed with discipline-specific methods and analysis software) are becoming increasingly relevant.

The aims of the present Methods Series launched by the Research Data Center PIAAC at GESIS are (a) to improve the accessibility of knowledge in keeping with the principle of open science (see, e.g., the Open Science Framework/OSF or the German Data Forum [RatSWD]); (b) to increase the sustainability of scientific work by documenting processes, thereby enabling them to be replicated by other researchers; and (c) to guide (early-career) researchers in their data analyses. Collecting and making available existing analysis scripts based on previous research with PIAAC data is one way of achieving these aims. A digital platform—[the PIAAC-scripts](#)—is being made available as an accompaniment to the Methods Series. It contains

analysis scripts for analyzing PIAAC data and other educational data in the adult research filed. This will facilitate the subsequent use of the analytical methods, and thus simplify and improve the scientific use of PIAAC data. Open access is an important asset to ensure transparency of data and methods, to give every researcher an equal chance, and to make replication studies easier. Only with sufficient insight into research data and methods is it possible to learn from each other and deliver correct results. The Methods Series establishes a guideline and provides examples of different methods and statistics programs (e.g., Stata, *Mplus*, and R). It also serves as quality assurance by providing simplified possibilities for replicating results and facilitating the reproduction of analyses.

The present chapter provides an overview of the available data sets and tools for the analysis of PIAAC data. In addition, it presents an example of differences in outcomes when using correct versus incorrect analyses procedures. Furthermore, it provides a brief introduction to PIAAC Discovery, a Shiny web application (app) for visualizing PIAAC data (Parker, n.d.). Chapter 2 to Chapter 4 present scripts from PIAAC workshops that took place online in 2022. These scripts show how to analyze PIAAC data with the software programs Stata ([Chapter 2](#)), R ([Chapter 3](#)), and *Mplus* ([Chapter 4](#)). And finally, Chapter 5 introduces scripts implemented in the International Database Analyzer (IDB Analyzer) to analyze PIAAC data.

If you are interested in having your own script published in the Methods Series, or in helping us to keep the [PIAAC scripts digital platform](#) up to date, please send us your research for suitability assessment. We would be happy to receive your submissions to support open data and to make it possible to present in this series the latest techniques for analyzing PIAAC data. The concept of openly sharing possible ways of analyzing PIAAC data is of great importance for future research, also in view of the new waves of the data sets coming in the future. It can lead to simpler, more accurate, and faster research results.

## 1.2 Availability of PIAAC Data

PIAAC is designed as a cross-sectional study to be repeated at regular intervals. Cycle 1 of PIAAC started in 2008 and comprised three rounds, in which a total of 38 countries participated (see [Table 1.1](#)). Round 1 took place in 2011–2012, and data were collected from a total of 24 countries. A further nine countries were added in Round 2 (2014–2015). In Round 3 of Cycle 1, in 2017, data for Ecuador, Hungary, Kazakhstan, Mexico, and Peru were collected. As Goldhammer et al. (2020) noted, PIAAC “was the first computer-based large-scale assessment to provide anonymised log file data from the cognitive assessment together with extensive online documentation and a data analysis support tool” (p. 239). PIAAC Cycle 1 provides log files from 17 countries, as well as scientific use files from four countries. The available data sets can be found on the [OECD](#) website or on the website of [the Research Data Center PIAAC](#) at GESIS.

An exemplary overview of data from PIAAC Germany and its longitudinal follow-up, PIAAC-L, and of how these data can be used, for example, in sociology or psychology can be found in Martin, Zabal et al. (2022) and Martin, Maehler et al. (2022).



**Table 1.1**

Overview of PIAAC Cycle 1 Countries and Available Data Sets

Country	Available data sets (provider)	Year of assessment (Cycle 1)
Australia	Public Use File (Australian Bureau of Statistics/OECD— <i>available only in the OECD version of the analysis tool PIAAC International Data Explorer [OECD IDE]</i> )	2011–2012
Austria	Public Use File (OECD) PIAAC Log Files (RDC PIAAC at GESIS) Scientific Use Files (Statistics Austria)	2011–2012 2011–2012 2011–2012
Belgium <sup>1</sup>	Public Use File (OECD) PIAAC Log Files (RDC PIAAC at GESIS)	2011–2012 2011–2012
Canada	Public Use File (OECD/Statistics Canada) Scientific Use File (Statistic Canada) Longitudinal and International Study of Adults (LISA) linked to register data (Canadian Research Data Centers)	2011–2012 2011–2012 2011–2012
Chile	Public Use File (OECD)	2014–2015
Cyprus	Public Use File (RDC PIAAC at GESIS)	2011–2012
Czech Republic	Public Use File (OECD)	2011–2012
Denmark	Public Use File (OECD) PIAAC Log Files (RDC PIAAC at GESIS) Nordic PIAAC Database linked to register data (Nordic NSIs <sup>2</sup> )	2011–2012 2011–2012 2011–2012
Ecuador	Public Use File (OECD)	2017
Estonia	Public Use File (OECD) PIAAC Log Files (RDC PIAAC at GESIS) Nordic PIAAC Database linked to register data (Nordic NSIs)	2011–2012 2011–2012 2011–2012
Finland	Public Use File (OECD) PIAAC Log Files (RDC PIAAC at GESIS) Nordic PIAAC Database linked to register data (Nordic NSIs)	2011–2012 2011–2012 2011–2012
France	Public Use File (OECD) PIAAC Log Files (RDC PIAAC at GESIS) PIAAC Data Files on Non-Cognitive Skills (RDC PIAAC at GESIS)	2011–2012 2011–2012 2017

Country	Available data sets (provider)	Year of assessment (Cycle 1)
Germany	Public Use File (OECD) PIAAC Log Files (RDC PIAAC at GESIS) Scientific Use Files (RDC PIAAC at GESIS) German PIAAC National Supplement: Prime Age (RDC PIAAC at GESIS) German PIAAC National Supplement: Competencies in Later Life/CiLL (RDC PIAAC at GESIS) PIAAC Germany Scientific Use File: Regional Data (RDC PIAAC at GESIS) PIAAC-L, Longitudinal Scientific Use File (RDC PIAAC at GESIS) PIAAC Data Files on Non-Cognitive Skills (RDC PIAAC at GESIS)	2011–2012 2011–2012 2011–2012 2011–2012 2011–2012 2011–2012 2012–2016 2017
Greece	Public Use File (OECD)	2014–2015
Hungary	Public Use File (OECD)	2017
Indonesia	<i>Not available</i>	2014–2015
Ireland	Public Use File (OECD) PIAAC Log Files (RDC PIAAC at GESIS)	2011–2012 2011–2012
Israel	Public Use File (OECD)	2014–2015
Italy	Public Use File (OECD) Public Use File – Extended (INAPP <sup>3</sup> ) PIAAC Log Files (RDC PIAAC at GESIS)	2011–2012 2011–2012 2011–2012
Japan	Public Use File (OECD) PIAAC Data Files on Non-Cognitive Skills (RDC PIAAC at GESIS)	2011–2012 2017
Kazakhstan	Public Use File (OECD)	2017
Korea	Public Use File (OECD) PIAAC Log Files (RDC PIAAC at GESIS)	2011–2012 2011–2012
Lithuania	Public Use File (OECD)	2014–2015
Mexico	Public Use File (OECD)	2017
Netherlands	Public Use File (OECD) PIAAC Log Files (RDC PIAAC at GESIS)	2011–2012 2011–2012
New Zealand	Public Use File (OECD) Public Use File – Extended (New Zealand Ministry of Education)	2014–2015 2014–2015

Country	Available data sets (provider)	Year of assessment (Cycle 1)
Norway	Public Use File (OECD) PIAAC Log Files (RDC PIAAC at GESIS) Norwegian PIAAC data linked to register data (NSD <sup>4</sup> ) Nordic PIAAC Database linked to register data (Nordic NSIs)	2011–2012 2011–2012 2011–2012 2011–2012
Peru	Public Use File (OECD)	2017
Poland	Public Use File (OECD) PIAAC Log Files (RDC PIAAC at GESIS) PIAAC Data Files on Non-Cognitive Skills (RDC PIAAC at GESIS)	2011–2012 2011–2012
Russian Federation	Public Use File (OECD)	2011–2012
Singapore	Public Use File (OECD)	2014–2015
Slovak Republic	Public Use File (OECD) PIAAC Log Files (RDC PIAAC at GESIS)	2011–2012 2011–2012
Slovenia	Public Use File (OECD)	2014–2015
Spain	Public Use File (OECD) PIAAC Log Files (RDC PIAAC at GESIS) PIAAC Data Files on Non-Cognitive Skills (RDC PIAAC at GESIS)	2011–2012 2011–2012 2017
Sweden	Public Use File (OECD) Swedish PIAAC data linked to register data (Statistics Sweden) Nordic PIAAC Database linked to register data (Nordic NSIs)	2011–2012 2011–2012 2011–2012
Turkey	Public Use File (OECD)	2014–2015
United Kingdom <sup>5</sup>	Public Use File (OECD) PIAAC Log Files (RDC PIAAC by GESIS) PIAAC Data Files on Non-Cognitive Skills (RDC PIAAC at GESIS)	2011–2012 2011–2012 2016

Country	Available data sets (provider)	Year of assessment (Cycle 1)
United States	Public Use File (OECD/NCES <sup>6</sup> ) PIAAC 2012/2014 US National Supplement Public Use Data File – Household (NCES) PIAAC Log Files (RDC PIAAC at GESIS) US PIAAC Restricted Use File (NCES) PIAAC 2012/2014 US National Supplement Restricted Use Data File – Household (NCES) PIAAC 2014 US National Supplement Public Use Data Files -Prison (NCES) PIAAC 2014 US National Supplement Restricted Use Data Files -Prison (NCES) PIAAC Data Files on Non-Cognitive Skills (RDC PIAAC at GESIS)	2011–2012 2011–2012; 2014 2011–2012; 2011–2012 2011–2012; 2014 2014 2014 2016

Note. Adapted from Maehler & Konradt (2020) and <https://www.oecd.org/skills/piaac/data/>.

<sup>1</sup> Only Flanders; <sup>2</sup> Nordic National Statistical Institutions – Denmark, Estonia, Finland, Norway, Sweden; <sup>3</sup> Istituto Nazionale per l'Analisi delle Politiche Pubbliche, Italy; <sup>4</sup> Norwegian Centre for Research Data, Norway; <sup>5</sup> Only England and Northern Ireland; <sup>6</sup> National Center for Education Statistics, United States.

### 1.3 PIAAC Methodology: Analyzing PIAAC Data

Table 1.2 provides an overview of existing tools and packages and helpful commands to analyze PIAAC data using different software programs (e.g., Stata, R, and *Mplus*). The tools that can be used to perform correct analyses using public use files (PUFs), scientific use files (SUFs), or log file data are: (a) the OECD version of the PIAAC International Data Explorer (OECD IDE) and the United States version of the PIAAC International Data Explorer (US IDE; for a description of the two versions, see Pawlowski & Soroui, 2020); (b) the International Association for the Evaluation of Educational Achievement (IEA) International Database Analyzer (IDB Analyzer; for a description, see Sandoval-Hernandez & Carrasco, 2020); and (c) the PIAAC Log Data Analyzer (LDA; for a description, see Goldhammer et al., 2020).

**Table 1.2**

Overview of Tools for the Analysis of PIAAC Data

Software program	Package	Source	Data used	Analysis/ Commands
<b>Stata</b>	PIAACTOOLS	Pokropek & Jakubowski (2019) Available at: <a href="https://www.oecd.org/skills/piaac/PIAC-TOOLS_16OCT_for_web.pdf">https://www.oecd.org/skills/piaac/PIAC-TOOLS_16OCT_for_web.pdf</a>	Public Use Files	Command <i>piacdes</i> for descriptive statistics (variance, mean, percentiles); <i>piactab</i> for tables and crosstables, and <i>piacreg</i> for linear and logistic regressions.
<b>Stata</b>	REPEST	Avvisati & Keslair (2014); Keslair (2020)	Public Use Files	Descriptive statistics (e.g., mean, variance, percentiles); tables and cross tables; regression analyses (e.g., linear, logistic)
<b>R</b>	EdSurvey	Bailey et al. (2020) See also: <a href="https://www.air.org/project/nces-data-r-project-edsurvey">https://www.air.org/project/nces-data-r-project-edsurvey</a>	Public Use Files	Descriptive statistics (mean, percentile analysis); gap analysis; proficiency levels; correlations, regression analyses (e.g., linear, logistic, multivariate)
<b>R</b>	intsvy	Caro & Biecek (2022) Available at: <a href="https://cran.r-project.org/web/packages/intsvy/intsvy.pdf">https://cran.r-project.org/web/packages/intsvy/intsvy.pdf</a> <a href="https://github.com/eldafani/intsvy">https://github.com/eldafani/intsvy</a>	Public Use Files	Descriptive statistics (mean, percentile analysis); proficiency levels; correlation, regression (e.g., logistic)
<b>R</b>	svyPVpack	Peterbauer & Reif (2014) Available at: <a href="https://github.com/manuelreif/svyPVpack">https://github.com/manuelreif/svyPVpack</a>	Public Use Files	Descriptive statistics (mean, percentile analysis); proficiency levels; correlation

Software program	Package	Source	Data used	Analysis/ Commands
<b>Mplus</b>		Scherer (2020) Available at: <a href="https://osf.io/hgbfk/">https://osf.io/hgbfk/</a>	Public Use Files	Confirmatory factor analysis (CFA); structural equation modeling (SEM; e.g., path models, multi-group versions)
<b>OECD PIAAC International Data Explorer (OECD IDE)</b>	Data Interface	Pawlowski & Soroui (2020) Available at: <a href="https://piaac-dataexplorer.oecd.org/ide/idepiaac/">https://piaac-dataexplorer.oecd.org/ide/idepiaac/</a>	Public Use Files	Descriptive statistics (mean, percentiles); proficiency levels
<b>US PIAAC International Data Explorer (US IDE)</b>	Data Interface	Pawlowski & Soroui (2020) Available at: <a href="https://nces.ed.gov/surveys/piaac/ideuspiaac/">https://nces.ed.gov/surveys/piaac/ideuspiaac/</a>	Public Use Files	Descriptive statistics (mean, percentiles); proficiency levels; gap analyses; linear regression
<b>IEA International Database Analyzer (IDB Analyzer)</b>	Windows-based tool with macro output in SPSS and SAS	Sandoval-Hernandez & Carrasco (2020); International Association for the Evaluation of Educational Achievement (IEA; n.d.) Available at: <a href="https://www.iea.nl/data-tools/tools#section-308">https://www.iea.nl/data-tools/tools#section-308</a>	Public Use Files; Scientific Use Files	Descriptive statistics (mean, percentiles); proficiency levels; correlations; regression analyses (logistic regression, linear regression)
<b>PIAAC LogData Analyzer (LDA)</b>	Data Interface	Goldhammer et al. (2020) Available at: <a href="https://piaac-logdata.tba-hosting.de/download/">https://piaac-logdata.tba-hosting.de/download/</a>	Public Use Files; PIAAC Log Files	Tool to extract predefined aggregated variables from XML files and raw log file data from XML files for creating user-defined aggregated variables.

As Arikan et al. (2020) noted, “international large-scale assessments ... play a key role in determining educational policies besides their primary objectives of measuring, evaluating and monitoring the educational process” (p. 43). They stressed that it was therefore “critical to analyze the data ... using scientifically accurate statistical methods as the results have the potential to influence millions of stakeholders through major policy changes” (p. 43).

In their study, Arikan et al. (2020) illustrated exemplarily what happens if basic methodological issues are not considered correctly when analyzing data from large-scale international assessments such as PIAAC. To raise awareness of the correct way to analyze these data, and to show that there is a high probability of getting incorrect results if sample weights and plausible values (PVs) are not used, the authors ran analyses with and without sample weights and plausible values.

Arikan et al. (2020) conducted *t*-test and multiple regression analyses using the IDB Analyzer (an interface from the International Data Explorer [IDE] for handling analyses of data from large-scale assessments such as PIAAC; for more information see Sandoval-Hernandez & Carrasco, 2020) and multilevel regression analysis using *Mplus*. For example, the results of the research question “Is there a statistically significant difference between mean PIAAC 2015 reading scores of adults who looked for a job last month and who did not look for a job last month in Turkey?” (Arikan et al., 2020, p. 51) revealed that when sample weights and plausible values were used, there was no statistically significant difference between the mean PIAAC 2015 literacy of adults in Turkey who looked for a job in the past month and those who did not (see results based on the IDB Analyzer in the first line of [Table 1.3](#)).

As [Table 1.3](#) shows, among the 11 analyses, there were contradictory results. The correct results in this case were those of the analyses based on the IDB Analyzer. Three of the analyses using only one PV showed significant differences—deviating from the correct result. The standard errors were higher when sample weights and PVs were used, than when this was not the case. As Arikan et al. (2020) noted, the change in the standard error directly affected the *t* value and the ultimate decision.

**Table 1.3**

*Exemplary Analyses of Literacy Among Adults in Turkey Who Looked for a Job in the Past Month and Those Who Did Not*

Method	Looked for a job (SE)	Did not look for a job (SE)	Mean difference (SE)	<i>t</i>
IDB Analyzer PV1–PV10	226.11 (4.16)	221.05 (1.45)	5.06 (4.36)	1.16
SPSS PV1	229.06 (2.51)	223.90 (.83)	5.16 (2.73)	1.89
SPSS PV2	229.40 (2.59)	223.23 (.83)	6.17 (2.75)	2.25*
SPSS PV3	227.01 (2.57)	224.33 (.83)	2.67 (2.73)	.98
SPSS PV4	226.87 (2.45)	224.12 (.84)	2.76 (2.74)	1.01
SPSS PV5	226.52 (2.55)	222.94 (.83)	3.58 (2.71)	1.32

Method	Looked for a job (SE)	Did not look for a job (SE)	Mean difference (SE)	t
SPSS PV6	231.42 (2.58)	224.81 (.84)	6.61 (2.75)	2.40*
SPSS PV7	226.62 (2.55)	223.93 (.82)	2.70 (2.71)	1.00
SPSS PV8	226.73 (2.56)	223.70 (.83)	3.03 (2.72)	1.11
SPSS PV9	227.88 (2.51)	222.49 (.84)	5.39 (2.76)	1.95
SPSS PV10	231.14 (2.63)	222.82 (.84)	8.32 (2.75)	3.02**
SPSS PVmean	228.27 (2.34)	223.63 (.76)	4.64 (2.51)	1.85

Note: Adapted from Arıkan et al. (2020, p. 52). PV = plausible value.

\* $p < .05$ . \*\* $p < .01$ . \*\*\*  $p < .001$ .

For a correct way to analyze the PIAAC data using different software and disciplinary approaches, see the methodological handbook by Maehler & Rammstedt (2020).

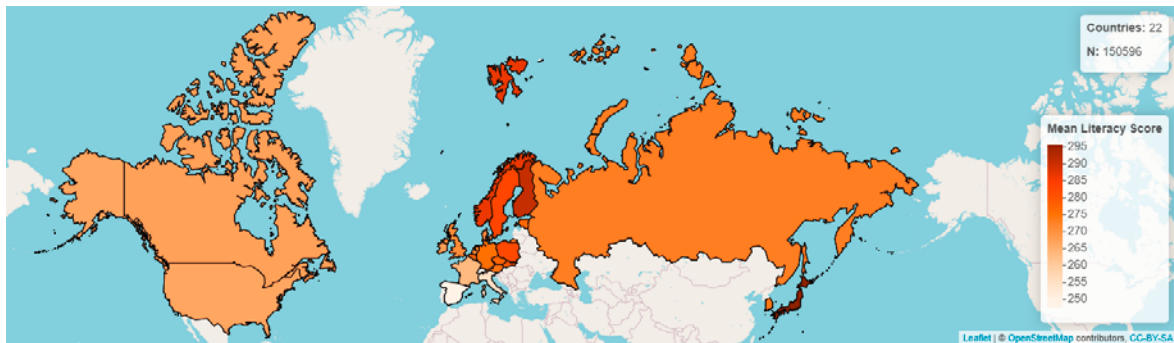
## 1.4 PIAAC Discovery: A Shiny App for Visualizing PIAAC Data

PIAAC Discovery is a Shiny app that offers a playful way to visualize and explore PIAAC data on an interactive map. Designed by Daniel R. Parker (n.d.), it allows a preliminary look at the numeracy and literacy skills scores of 22 countries that participated in Cycle 1, Round 1 of PIAAC (for a list of these countries, see [Table 1.1](#)).

Users of the app can apply different filters (e.g., age, gender, parents' education, earnings) to take variables into account. By clicking on the individual countries, the mean score is displayed. Countries and sample size are displayed on the right-hand side of the map (see the exemplary map in [Figure 1.1](#)).

Parker (n.d.) stresses that “the application is not associated with PIAAC or the OECD,” and that its purpose “is to guide research interest rather than test hypotheses” (Information page). To date, only one PV has been taken into consideration, which makes the results unreliable. Additionally, the data have not been weighted for analysis, and “any standard errors produced from the data are biased” (Parker, n.d., Information page).





**Figure 1.1** Literacy by country displayed by the Shiny app PIAAC Discovery (Parker, n.d.)

Source of underlying data: <https://www.oecd.org/skills/piaac/data/>

## 1.5 References

- Arıkan, S., Özer, F., Şeker, V. & Ertaş, G. (2020). The importance of sample weights and plausible values in large-scale assessments. *Journal of Measurement and Evaluation in Education and Psychology*, 11(1), 43–60. <https://doi.org/10.21031/epod.602765>
- Avvisati, F., & Keslair, F. (2014). *REPEST: Stata module to run estimations with weighted replicate samples and plausible values* (Revised January 6, 2020) [Computer software]. Boston College, Department of Economics. <https://econpapers.repec.org/software/bocbocode/s457918.htm>
- Bailey, P., Lee, M., Nguyen, T., & Zhang, T. (2020). Using EdSurvey to analyse PIAAC data. In D. B. Maehler & B. Rammstedt (Eds.), *Large-scale cognitive assessment* (pp. 209–237). Springer. [https://doi.org/10.1007/978-3-030-47515-4\\_9](https://doi.org/10.1007/978-3-030-47515-4_9)
- Caro, D., & Biecek, P. (2022). *Package ‘intsvy’* [Computer software]. CRAN. <https://cran.r-project.org/web/packages/intsvy/intsvy.pdf>
- Gal, I., Alatorre, S., Close, S., Evans, J., Johansen, L., Maguire, T., Manly, M., & Tout, D. (2009). *PIAAC numeracy: A conceptual framework* (OECD Education Working Papers, No. 35). OECD Publishing. <https://doi.org/10.1787/220337421165>
- Goldhammer, F., Hahnel, C., & Kroehne, U. (2020). Analysing log file data from PIAAC. In D. B. Maehler & B. Rammstedt (Eds.), *Large-scale cognitive assessment* (pp. 239–269). Springer. [https://doi.org/10.1007/978-3-030-47515-4\\_10](https://doi.org/10.1007/978-3-030-47515-4_10)
- International Association for the Evaluation of Educational Achievement (IEA). (n.d.). *IDB Analyzer* [Computer software]. IEA. <https://www.iea.nl/data-tools/tools#section-308>
- Keslair, F. (2020). Analysing PIAAC data with Stata. In D. B. Maehler & B. Rammstedt (Eds.), *Large-scale cognitive assessment* (pp. 149–164). Springer. [https://doi.org/10.1007/978-3-030-47515-4\\_7](https://doi.org/10.1007/978-3-030-47515-4_7)
- Jones, S., Gabrielsen, E., Hagston, J., Linnakylä, P., Megherbi, H., Sabatini, J., Tröster, M., & Vidal-Abarca, E. (2009). *PIAAC literacy: A conceptual framework* (OECD Education Working Papers, No. 34). OECD Publishing. <http://doi.org/10.1787/220348414075>
- Maehler, D. B., & Konradt, I. (2022). *PIAAC bibliography 2008–2021* (GESIS Papers, 2022/02). GESIS – Leibniz Institute for the Social Sciences. <https://doi.org/10.21241/ssoar.77833>
- Maehler, D. B., & Konradt, I. (2020). Adult cognitive and non-cognitive skills: An overview of existing PIAAC data. In D. B. Maehler & B. Rammstedt (Eds.), *Large-scale cognitive assessment* (pp. 49–92). Springer. [https://doi.org/10.1007/978-3-030-47515-4\\_4](https://doi.org/10.1007/978-3-030-47515-4_4)
- Maehler, D. B., & Rammstedt, B. (Eds.). (2020). *Large-scale cognitive assessment*. Springer.

- Martin, S., Maehler, D. B., Zabal, A., & Rammstedt, B. (2022). PIAAC-L: The longitudinal follow-up to PIAAC in Germany. *Soziale Welt – Zeitschrift für sozialwissenschaftliche Forschung und Praxis*, 73(1), 169–199. <https://doi.org/10.5771/0038-6073-2022-1-169>
- Martin, S., Zabal, A., Maehler, D. B., & Rammstedt, B. (2022). Data from PIAAC Germany and its longitudinal follow-up, PIAAC-L. *Journal of Open Psychology Data*, 10(1), 20. <https://doi.org/10.5334/jopd.74>
- Organisation for Economic Co-operation and Development (OECD). (2013). *OECD skills outlook 2013: First results from the Survey of Adult Skills*. OECD Publishing. <http://dx.doi.org/10.1787/9789264204256-en>
- Organisation for Economic Co-operation and Development (OECD). (2021). *The assessment frameworks for Cycle 2 of the Programme for the International Assessment of Adult Competencies*. OECD Publishing. <https://doi.org/10.1787/4bc2342d-en>
- Parker, D. R. (n.d.). *PIAAC Discovery* [Web application]. [https://danielallenparker.shinyapps.io/PIAAC\\_Discovery/](https://danielallenparker.shinyapps.io/PIAAC_Discovery/)
- Pawlowski, E., & Soroui, J. (2020). Analysing PIAAC data with the International Data Explorer (IDE). In D. B. Maehler & B. Rammstedt (Eds), *Large-scale cognitive assessment* (pp. 93–115). Springer. [https://doi.org/10.1007/978-3-030-47515-4\\_5](https://doi.org/10.1007/978-3-030-47515-4_5)
- Peterbauer, J., & Reif, M. (2014). *svyPVpack* [Computer software]. GitHub. <https://github.com/manuelreif/svyPVpack>
- Pokropek, A., & Jakubowski, M. (2019). *PIACTOOLS: Stata® programs for statistical computing using PIAAC data*. OECD Publishing. [https://www.oecd.org/skills/piaac/PIACTOOLS\\_16OCT\\_for\\_web.pdf](https://www.oecd.org/skills/piaac/PIACTOOLS_16OCT_for_web.pdf)
- Rouet, J.-F., Bétrancourt, M., Britt, M. A., Bromme, R., Graesser, A. C., Kulikowich, J. M., Leu, D. J., Ueno, N., & van Oostendorp, H. (2009). *PIAAC problem solving in technology-rich environments: A conceptual framework* (OECD Education Working Papers, No. 36). OECD Publishing. <http://dx.doi.org/10.1787/220262483674>
- Sandoval-Hernandez, A., & Carrasco, D. (2020). Analysing PIAAC data with the IDB Analyzer (SPSS and SAS). In D. B. Maehler & B. Rammstedt (Eds), *Large-scale cognitive assessment* (pp. 117–148). Springer. [https://doi.org/10.1007/978-3-030-47515-4\\_6](https://doi.org/10.1007/978-3-030-47515-4_6)
- Scherer, R. (2019, June 21). *PIAAC structural equation modeling with Mplus* [Project]. <https://doi.org/10.17605/OSF.IO/HGBFK>

## 2 Analyzing PIAAC data with Stata

*Author: Britta Gauly*

*(GESIS – Leibniz-Institute for the Social Sciences, Germany)*

### 2.1 Introduction

The scripts presented in this section are based on the [online tutorial](#) “PIAAC Data Analysis in Stata”. They provide a practical guide to performing simple analyses of data from the Programme for the International Assessment of Adult Competencies (PIAAC) in Stata. The tutorial is available on the website of the Research Data Center PIAAC ([RDC PIAAC](#)) at GESIS (RDC PIAAC 2021a, 2021b, 2021c, 2021d). The target audience of the tutorial are researchers with at least some experience in data analysis in Stata, who have little or no experience with the analysis of PIAAC data, and who are interested in finding out whether the PIAAC data are suitable for answering their research questions. The tutorial provides several simple example analysis tasks that can be performed with PIAAC data.

[Section 2.2](#) provides information on how to obtain access to the international PIAAC data. [Sections 2.3](#) and [2.4](#) show how different analysis tools work, what types of analyses they provide, and the advantages and disadvantages of these tools.

### 2.2 Data Access and First Steps in Stata

The public use files (PUFs) for most countries participating in PIAAC can be accessed via the OECD’s “PIAAC Data” [webpage](#). As these data are not in Stata format (.dta) by default, they have to be converted. By clicking on “Download the datasets (Public Use Files)” under the header “PIAAC Data,” you can download the data in SAS, SPSS, or CSV format.

First, you have to download the data set for each country in CSV format (see OECD, 2016b to OECD, 2016jj), and save all files in one folder on your computer. In the second step, you have to download the do-file in order to import the CSV data into Stata and save it on your computer. This do-file will import and append all PUFs into a unique Stata data set. Third, you can run the do-file by clicking on the “Execute” or “Do” button in the upper left-hand corner. While running, Stata will request the file path where you saved all CSV files. Type the path in the command window and click on “Enter.” You will receive the message: “Your PIAAC dataset is now ready to be used in Stata. Please consider saving this file.” To save the file, type “save” and the path where you want to store your data and a file name in the command window, and click on “Enter.”

This procedure will provide you with the PIAAC PUFs, which are freely available on the OECD website. As data may be updated, it is worth checking the OECD website from time to time.

The suggested folder structure for the data analysis presented below is as follows:

- `orig` includes the original data, comprising the CSV files and the original Stata data set you received when you ran the do-file;
- `data` contains all user-generated data sets, for example, comprising only a subset of countries or respondents;
- `log` includes all log files produced during data analysis;
- `prog` includes all do files;
- `out` contains the result files.

```

/* Load data with all countries */
use "${orig}PIAAC_All_PUFs.dta"

/* Tabulate countries in the data set */
tab centryid

/* Some countries do not have a label yet --> change label of variable
"centryid" */
label define CENTRYID 152 "Chile" 218 "Ecuador" ///
300 "Greece" 348 "Hungary" 376 "Indonesia" ///
398 "Kazakhstan" 440 "Lithuania" ///
484 "Mexico" 554 "New Zealand" 604 "Peru" ///
702 "Singapore" 705 "Slovenia" 792 "Turkey", add
label val centryid CENTRYID

tab centryid

/* To get a first look at the skill variables (plausible values), type pvlit,
pvnum, pvpsl in the variable search window or display summary statistics */
sum pvlit1

/* To get a first look at the weighting variables, type spfwt in the variable
search window or display summary statistics */

/* Save small data set including Germany, Greece, Singapore, Sweden, US --> we
will use this data set in the following sample analyses */
preserve
keep if centryid == 276 | centryid == 300 | centryid == 702 | centryid == 752 |
centryid == 840
save "${data}PIAAC_5Countries.dta", replace
restore

```

## 2.3 Analyzing PIAAC Data Using the Package PIACTOOLS

The package PIACTOOLS was developed by Jakubowski and Pokropek (2019).

```

/* Install piaaltools */
ssc install piaactools, replace

/* Help function */
help piaacdes
help piaactab
help piaacreg

/* Define global paths */
global data "...\\2_data\\"
global log "...\\3_log\\"
global out "...\\5_out\\"

/* Use log file */
log using "${log}Examples_piaactools.log", replace

/* Load data */
use "${data}PIAAC_5Countries.dta"

/* Tabulate countries in the dataset */
tab centryid

```

### 2.3.1 Examples *piaacdes*

```

/* Example 1: Average years of education */
piaacdes yrsqual, countryid(centryid) stats(mean) round(2) save("${out}
piaacdes_Ex1")

/* Example 2: Average literacy skills, overall and by gender */
piaacdes, pv(pvlit) countryid(centryid) stats(mean) round(2) save("${out}
piaacdes_Ex2a")

piaacdes, pv(pvlit) countryid(centryid) stats(mean) over(gender_r) round(2)
save("${out}piaacdes_Ex2b")

/* Example 3: Literacy skill dispersion (5th, 25th, 75th, 95th quantile);
overall and for men between 16 and 34 years */
piaacdes, pv(pvlit) countryid(centryid) centile(5 25 75 95) round(2)
save("${out}piaacdes_Ex3a")

piaacdes if gender_r == 1 & ageg10lfs < 3, pv(pvlit) countryid(centryid)
centile(5 25 75 95) round(2) save("${out}piaacdes_Ex3b")

```

### 2.3.2 Examples *piaactab*

```

/* Example 1: Percentages of educational qualifications of respondents'
mothers */
piaactab j_q06b, countryid(cntryid) round(1) save("${out}piaactab_Ex1")

/* Example 2: Percentages of respondents at each numeracy level, overall and
for the employed population */
piaactab pvnum, countryid(cntryid) round(1) save("${out}piaactab_Ex2a")

piaactab pvnum if c_d05 == 1, countryid(cntryid) round(1) save("${out}
piaactab_Ex2b")

/* Example 3: Crosstable of numeracy skills and native language */
piaactab pvnum, over(nativelang) countryid(cntryid) round(1) save("${out}
piaactab_Ex3")

/* Declare missing values in variable "nativelang" */
tab nativelang
mvdecode nativelang, mv(8, 9)

piaactab pvnum, over(nativelang) countryid(cntryid) round(1) save("${out}
piaactab_Ex3new")

```

### 2.3.3 Examples *piaacreg*

```

/* Example 1: How are age, gender, formal education, and computer experience
in the job related to PS-TRE skills? */

/* Recode the variable "computer experience" to ease interpretation */
recode g_q04 (1=1) (2=0)

/* Create dummy variables for the categorical variables "age" and "education"
and include all but the reference categories (here: youngest age group and low
education) to the regression */
tab ageg10lfs, gen(age10dum)
tab1 age10dum*

tab edcat6, gen(ed6dum)
tab1 ed6dum*

piaacreg age10dum2 age10dum3 age10dum4 age10dum5 gender_r ed6dum2 ed6dum3
ed6dum4 ed6dum5 ed6dum6 g_q04, pvdep(pvpsl) countryid(cntryid) round(2)
save("${out}piaacreg_Ex1b")

/* Example 2: Do literacy skills and formal education determine participation
in adult education for women between 35 and 54 years? */

/* First step: check whether the dependent variable is a binary variable */

```

```

tab nfe12

piaacreg nfe12 ed6dum2 ed6dum3 ed6dum4 ed6dum5 ed6dum6 , pvindep1(pvlit)
cmd("logit") countryid(cntryid) round(2) save("${out}piaacreg_Ex2a")

/* The coefficient on literacy is rather small as literacy is measured on a
scale from 0-500. Standardize literacy skills to get the relationship between
a one standard deviation increase in literacy skills and participation in
training */
forv i = 1/10 {
qui egen pvlit_std`i' = std(pvlit`i')
}

piaacreg nfe12 ed6dum2 ed6dum3 ed6dum4 ed6dum5 ed6dum6 , pvindep1(pvlit_std)
cmd("logit") countryid(cntryid) round(2) save("${out}piaacreg_Ex2b")

log close

```

## 2.4 Analyzing PIAAC Data Using the REPEST Macro

The REPEST macro was developed by Avvisati and Keslair (2014; Keslair, 2020).

```

/* Install piaaltools */
ssc install repest, replace

/* Help function */
help repest

/* Define global paths */
global data "...\\2_data\\"
global log "...\\3_log\\"
global out "...\\5_out\\"

/* Use log file */
log using "${log}Examples_repest.log", replace

/* Load data */
use "${data}PIAAC_5Countries.dta"

/* Tabulate countries in the dataset */
tab cntryid

```

### 2.4.1 Examples Descriptive Statistics

```

/* Example 1: Average years of education and literacy skills; overall and by
gender */

/* The @ character indicates variables with plausible values and initiates a
loop over the correct number of plausible values (here: 10) */

```

```

repest PIAAC, estimate(means yrsqual pvlit@) by(cntryid, average(276 300 702
752 840))

repest PIAAC, estimate(means yrsqual pvlit@) by(cntryid, average(276 300 702
752 840)) over(gender_r)

/* Example 2: Literacy skill dispersion (5th, 25th, 75th, 95th quantile);
overall and for men between 16 and 34 years */
repest PIAAC, estimate(summarize pvlit@, stats(p5 p25 p75 p90)) by(cntryid,
levels(276 752))

repest PIAAC if gender_r == 1 & ageg10lfs == 1, estimate(summarize pvlit@,
stats(p5 p25 p75 p90)) by(cntryid, levels(276 752))

/* Example 3: Percentages of respondents at each numeracy level, overall and
for the employed population */

/* Create levels of numeracy skills according to the OECD (2016a, p.71)
definition */
forv i = 1/10 {
    gen          numlevel`i' = .
    qui replace numlevel`i' = 1 if pvnum`i' <= 225
    qui replace numlevel`i' = 2 if pvnum`i' > 225 & pvnum`i' <= 275
    qui replace numlevel`i' = 3 if pvnum`i' > 275 & pvnum`i' <= 325
    qui replace numlevel`i' = 4 if pvnum`i' > 325
    qui replace numlevel`i' = . if pvnum`i' == .
}

repest PIAAC, estimate(freq numlevel@) by(cntryid) outfile("${out}
repest_Ex3a")

repest PIAAC if c_d05 == 1, estimate(freq numlevel@) by(cntryid)
outfile("${out}repest_Ex3b")

/* Example 4: Crosstable of numeracy skills and native language */

/* Declare missing values in variable "nativelang" */
tab nativelang
mvdecode nativelang, mv(8, 9)

repest PIAAC, estimate(freq numlevel@) over(nativelang)
repest PIAAC, estimate(means pvnum@) over(nativelang, test)

/* Example 5: Correlations between literacy, numeracy, and problem-solving
skills */
repest PIAAC, estimate(corr pvlit@ pvnum@ pvpsl@) by(cntryid, average(276 300
702 752 840))

```



## 2.4.2 Examples Regression Analysis

```

/* Example 1: How are age, gender, formal education, and computer experience
in the job related to PS-TRE skills? */

/* Recode the variable "computer experience" to ease interpretation */
recode g_q04 (1=1) (2=0)

repest PIAAC, estimate(stata: reg pvps1@ i.ageg101fs gender_r i.edcat6 g_q04)
by(cntryid) results(add(N r2))

/* Default: The youngest age group and the lowest education category are the
reference groups */

/* Example 2: Do literacy skills and formal education determine participation
in adult education for women between 35 and 54 years? */

/* First step: check whether the dependent variable is a binary variable */
tab nfe12

repest PIAAC, estimate(stata: logistic nfe12 i.edcat6 pvlit@) by(cntryid,
levels(276 840)) results(add(N r2))

log close

```

## 2.5 References

- Avvisati, F. & Keslair, F. (2014). *REPEST: Stata module to run estimations with weighted replicate samples and plausible values* (Revised January 6, 2020) [Computer software]. Boston College Department of Economics. <https://econpapers.repec.org/software/bocbocode/s457918.htm>
- Jakubowski, M., & Pokropek, A. (2019). PIAACTOOLS: A program for data analysis with PIAAC data. *The Stata Journal*, 19(1), 112–128. <https://doi.org/10.1177/1536867X19830909>
- Keslair, F. (2020). Analysing PIAAC data with Stata. In D. B. Maehler & B. Rammstedt (Eds.), *Large-scale cognitive assessment* (pp. 149–164). Springer. [https://doi.org/10.1007/978-3-030-47515-4\\_7](https://doi.org/10.1007/978-3-030-47515-4_7)
- Organisation for Economic Cooperation and Development (OECD). (2016a). *The Survey of Adult Skills: Reader's companion* (2nd ed.) OECD Publishing. <http://dx.doi.org/10.1787/9789264258075-en>
- Organisation for Economic Cooperation and Development (OECD). (2016b). *Programme for the International Assessment of Adult Competencies (PIAAC), Austria Public Use File* [Version: 17343010, prgautp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016c). *Programme for the International Assessment of Adult Competencies (PIAAC), Belgium Public Use File* [Version: 18224205, prgbelp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016d). *Programme for the International Assessment of Adult Competencies (PIAAC), Canada Public Use File* [Version: 88830378, prgcanp1.csv]. OECD Publishing.

- Organisation for Economic Cooperation and Development (OECD). (2016e). *Programme for the International Assessment of Adult Competencies (PIAAC), Chile Public Use File* [Version: 17135698, prgchlp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016f). *Programme for the International Assessment of Adult Competencies (PIAAC), Czech Republic Public Use File* [Version: 20736629, prgczep1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016g). *Programme for the International Assessment of Adult Competencies (PIAAC), Denmark Public Use File* [Version: 24972525, prgdnkp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016h). *Programme for the International Assessment of Adult Competencies (PIAAC), Ecuador Public Use File* [Version: 16788347, prgecup1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016i). *Programme for the International Assessment of Adult Competencies (PIAAC), Estonia Public Use File* [Version: 25276973, prgestp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016j). *Programme for the International Assessment of Adult Competencies (PIAAC), Finland Public Use File* [Version: 18842845, prgfinp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016k). *Programme for the International Assessment of Adult Competencies (PIAAC), France Public Use File* [Version: 23516989, prgrap1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016l). *Programme for the International Assessment of Adult Competencies (PIAAC), Germany Public Use File* (Version: 18834098, prgdeup1.csv). OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016m). *Programme for the International Assessment of Adult Competencies (PIAAC), Greece Public Use File* [Version: 15965250, prggrcp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016n). *Programme for the International Assessment of Adult Competencies (PIAAC), Hungary Public Use File* [Version: 20439451, prghunp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016o). *Programme for the International Assessment of Adult Competencies (PIAAC), Ireland Public Use File* [Version: 19982813, prgirlp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016p). *Programme for the International Assessment of Adult Competencies (PIAAC), Israel Public Use File* [Version: 18069090, prgisrp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016q). *Programme for the International Assessment of Adult Competencies (PIAAC), Italy Public Use File* [Version: 15433181, prgitap1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development ((OECD). (2016r). *Programme for the International Assessment of Adult Competencies (PIAAC), Japan Public Use File* [Version: 17505957, prgjnp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016s). *Programme for the International Assessment of Adult Competencies (PIAAC), Kazakhstan Public Use File* [Version: 20450859, prgkazp1.csv]. OECD Publishing.

- Organisation for Economic Cooperation and Development (OECD). (2016t). *Programme for the International Assessment of Adult Competencies (PIAAC), Korea Public Use File* [Version: 22217045, prgkorp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016u). *Programme for the International Assessment of Adult Competencies (PIAAC), Lithuania Public Use File* [Version: 17305986, prgltp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016v). *Programme for the International Assessment of Adult Competencies (PIAAC), Mexico Public Use File* [Version: 19760643, prgmexp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016w). *Programme for the International Assessment of Adult Competencies (PIAAC), Netherlands Public Use File* [Version: 18028845, prgnldp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016x). *Programme for the International Assessment of Adult Competencies (PIAAC), New Zealand Public Use File* [Version: 21235362, prgnzlp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016y). *Programme for the International Assessment of Adult Competencies (PIAAC), Norway Public Use File* [Version: 17723269, prgnorp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016z). *Programme for the International Assessment of Adult Competencies (PIAAC), Peru Public Use File* [Version: 22883339, prgperp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016aa). *Programme for the International Assessment of Adult Competencies (PIAAC), Poland Public Use File* [Version: 30634733, prgpolp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016bb). *Programme for the International Assessment of Adult Competencies (PIAAC), Russian Federation Public Use File* [Version: 13378965, prgrusp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016cc). *Programme for the International Assessment of Adult Competencies (PIAAC), Singapore Public Use File* [Version: 18353722, prgsgpp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016dd). *Programme for the International Assessment of Adult Competencies (PIAAC), Slovak Republic Public Use File* [Version: 18921861, prgsvkp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016ee). *Programme for the International Assessment of Adult Competencies (PIAAC), Slovenia Public Use File* [Version: 18125930, prgsvnp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016ff). *Programme for the International Assessment of Adult Competencies (PIAAC), Spain Public Use File* [Version: 20201797, prgespp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016gg). *Programme for the International Assessment of Adult Competencies (PIAAC), Sweden Public Use File* [Version: 15716978, prgswep1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016hh). *Programme for the International Assessment of Adult Competencies (PIAAC), Turkey Public Use File* [Version: 16765802, prgturp1.csv]. OECD Publishing.

- Organisation for Economic Cooperation and Development (OECD). (2016ii). *Programme for the International Assessment of Adult Competencies (PIAAC), United Kingdom Public Use File* [Version: 30110173, prggbp1.csv]. OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD). (2016jj). *Programme for the International Assessment of Adult Competencies (PIAAC), United States Public Use File* [Version: 17016050, prgusap1\_2012.csv]. OECD Publishing.
- Research Data Center PIAAC (RDC PIAAC). (2021a). *PIAAC data analysis in Stata: A practical guide. Online tutorial, Video 1 – Introduction to PIAAC data*. GESIS –Leibniz Institute for the Social Sciences. <https://www.gesis.org/en/piaac/rdc/>
- Research Data Center PIAAC (RDC PIAAC). (2021b). *PIAAC data analysis in Stata: A practical guide Online tutorial, Video 2 – Data access and first steps in Stata*. GESIS –Leibniz Institute for the Social Sciences. <https://www.gesis.org/en/piaac/rdc/>
- Research Data Center PIAAC (RDC PIAAC). (2021c). *PIAAC data analysis in Stata: A practical guide. Online tutorial, Video 3 – Stata-Ado piaactools*. GESIS –Leibniz Institute for the Social Sciences. <https://www.gesis.org/en/piaac/rdc/>
- Research Data Center PIAAC (2021d). *PIAAC data analysis in Stata: A practical guide. Online tutorial, Video 4 – Stata-Ado repest macro*. GESIS –Leibniz Institute for the Social Sciences. <https://www.gesis.org/en/piaac/rdc/>

## 3 Analyzing PIAAC Data With R

*Authors: Ting Zhang, Paul Bailey and Emmanuel Sikali  
(American Institutes for Research, Arlington, VA, United States)*

### 3.1 Introduction

This section explains how to use EdSurvey, an R package for large-scale assessment data, to analyze data from the Programme for the International Assessment of Adult Competencies (PIAAC). To analyze these data efficiently, taking into account their complex sample survey design and use of plausible values, the package uses procedures that follow PIAAC methodology.

The following subsections present scripts for essential PIAAC data analysis, from loading the data to data preparation, descriptive statistics, and statistical models (for detailed information see the [PIAAC scripts digital platform](#)).

### 3.2 Data Preparation and First Steps in R

#### 3.2.1 Loading and Reading in Data

This example shows how to download PIAAC Cycle 1 data all at once and use the `readPIAAC` command, which enables loading selected countries' PIAAC data into R as an `edsurvey.data.frame` (e.g., ita; OECD, 2016). To help users manage their system memory efficiently, EdSurvey loads data when needed, and otherwise does not use the memory. Only some metadata and results are maintained in the computer's random-access memory, including the file layout, missing value and special response codes, and assessment design attributes such as plausible values, achievement levels, weights, the primary sampling unit, and stratum variables.

```
library(EdSurvey)
# download data from the first cycle
downloadPIAAC(root = "~", cycle = 1)
# to read a selection of countries (e.g., Italy)
ita <- readPIAAC("~/PIAAC/Cycle 1/", countries = 'ITA')
```

To explore survey design attributes from an `edsurvey.data.frame`:

```
showPlausibleValues(ita, verbose = TRUE)
showWeights(ita)
getAttributes(ita, 'omittedLevels')
# by default, EdSurvey will show results from the analyses after listwise
# deletion of respondents with any special values, which are referred as
# 'omitted levels' in EdSurvey; for any data, the omitted levels can be seen
# with the omittedLevels command
ita # see all this information at once
```

### 3.2.2 Data Exploration Using EdSurvey

EdSurvey provides functions for data exploration and variables searching, including `showCodebook`, `searchSDF`, `levelsSDF`.

```
# show codebook
View(showCodebook(ita))

# search variables
searchSDF('income', data = ita)
searchSDF(c('income', 'annual'), data = ita) # search for an interview
question that includes given words
searchSDF(c('income', 'annual'), data = ita, levels=TRUE) # same + look at the
answers

# display variable levels
levelsSDF('d_q18a_t', data = ita)
```

### 3.3 Descriptive Statistics

The `summary2` function produces both weighted and unweighted descriptive statistics for a variable (e.g., 'd\_q18a\_t').

```
summary2(ita, 'd_q18a_t') # produce weighted for a variable "ANNUAL NET INCOME"
summary2(ita, 'd_q18a_t', weightVar = NULL) # produce unweighted descriptive
statistics
summary2(ita, "d_q18a_t", omittedLevels = TRUE) # exclude missing
summary2(ita, 'lit')
```

`edsurveyTable` creates a summary table of an outcome variable by selected categorical variables.

```
edsurveyTable(lit ~ ageg10lfs, data = ita) # table with literacy score by age
group
edsurveyTable(lit ~ ageg10lfs + c_d05, data = ita) # table with literacy score
by age and employment status
```

In its default setting, EdSurvey applies the sampling weights (e.g., 'spfw0'), the 10 plausible values of every scale or subscale of interest, and the jackknife method to compute the variance. It deals with special response codes and missing values by using listwise deletion. The reader can change these default settings using the list of arguments in the EdSurvey documentation.

In general, to find more information about any EdSurvey command, use "?command"; for example:

```
?edsurveyTable
```

### 3.3.1 Correlation

EdSurvey features multiple correlation methods for data exploration and analysis that fully account for the complex sample design of PIAAC by using the `cor.sdf` function. The following example displays how to conduct weighted polyserial correction between the literacy scale scores and “ANNUAL NET INCOME”.

```
cor.sdf('lit', 'd_q18a_t', data = ita, method = 'Polyserial')
```

## 3.4 Data Manipulation

EdSurvey allows for rudimentary data manipulation and analysis with both EdSurvey and base R functions to edit data before processing. Below are examples for some essential data manipulations using EdSurvey. See the [EdSurvey User Guide](#) (Lee et al., 2022) for more advanced data manipulations.

### Subset

```
itaM <- subset(ita, gender_r %in% 'MALE') # subset males
summary2(itaM, 'lit') # look for the literacy levels for males
```

### Recode

```
summary2(ita, 'c_d05') # look at the employment variable
ita$c_d05_recode <- ifelse(ita$c_d05 %in% c('OUT OF THE LABOUR FORCE',
                                           'UNEMPLOYED'), 'NOT EMPLOYED',
                           as.character(ita$c_d05))
edsurveyTable(lit ~ c_d05_recode, data = ita) # table literacy by recoded
employment status
```

### Rename

```
ita$emp <- ita$c_d05_recode
edsurveyTable(lit ~ emp, data = ita)
itaRaw <- getData(data = ita, varnames = c('lit', 'spfw0', 'gender_r', 'c_
d05'), addAttributes = FALSE, omittedLevels = TRUE)
# extract two variables: gender and annual income + 10 plausible values
associated with lit + the weight for this data frame: spfw0

itaRaw[1:5,1:15]
itaRaw$c_d05 <- gsub(pattern = 'OUT OF THE LABOUR FORCE|UNEMPLOYED',
                    replacement = 'not employed',
                    x = itaRaw$c_d05)
# gsub (base R function) uses pattern matching to replace values in a
variable, recodes the values in the variable c_d05
itaRaw <- subset(itaRaw, !c_d05 %in% 'NOT KNOWN') # removes the level 'NOT
KNOWN'
itaRawRebinded <- rebindAttributes(itaRaw, ita) # reassign survey attributes
for EdSurvey package
edsurveyTable(lit ~ c_d05, data = itaRawRebinded) # now we can apply EdSurvey
functions on the cleaned data
```

### 3.5 Data Analysis

EdSurvey offers various analytical functions that take account of the complex design of PIAAC data and its plausible values. This section showcases five of these functions: linear regression, logistic regression, gap analysis, percentile analysis, and proficiency level analysis. See the [EdSurvey User Guide](#) (Lee et al., 2022) for more analytical functions, including mixed models, multivariate regression, and quantile regression.

#### 3.5.1 Linear Regression

The data are read in and analyzed by the `lm.sdf` function. In the example, income quintile and age are regressed on the 10 plausible values for the literacy scale. By default, full sample weight “`spfw0`” is applied, and variance is estimated using the jackknife method.

```
lm1 <- lm.sdf(lit ~ d_q18a_t + age_r, data = ita)
# literacy score (lit), which is described as a function of income quintile
# (d_q18a_t) and age (age_r)
summary(lm1)
# note that there is no need to generate dummy codes for discrete variables
# like d_q18a_t
```

#### 3.5.2 Logistic Regression

The `logit.sdf` function predicts binary outcomes from a set of predictors. Although some variables might already be binary, the function `I()` and a rule such as `lit > 300` will dichotomize a nonbinary variable and specify the desired outcome level. The following example dichotomizes the literacy with the level over 300 as the outcome level. The outcome variable could also be a contextual variable.

```
logit1 <- logit.sdf(I(lit > 300) ~ ageg101fs + j_q06b, data = ita)
summary(logit1)$coefmat
# omitted values are excluded
oddsRatio(logit1) # works only for results from the logit.sdf function
waldTest(model = logit1, coef = 'j_q06b')
# this is a test of both coefficients in j_q06b being zero; two test results
# are shown: the chi-square test and the F-test
```

#### 3.5.3 Gap Analysis

In the following example, using the `gap` function, we compare literacy scores of self-employed persons with those of employees.

```
gap(variable = 'lit', data = ita, groupA = d_q04 %in% 'SELF-EMPLOYED',
     groupB = d_q04 %in% 'EMPLOYEE')
```



### 3.5.4 Percentile Analysis

The `percentile` function compares a numeric vector of percentiles in the range 0 to 100 for a data year. This function can display percentiles of scores for a selected group (e.g., males or females) in the examples below.

```
percentile(variable = 'lit', percentiles = c(10, 25, 50, 75, 90), data = ita)
percentile(variable = 'lit', percentiles = c(25, 50, 75), data = subset(ita,
gender_r %in% 'MALE'))
percentile(variable = 'lit', percentiles = c(25, 50, 75), data = subset(ita,
gender_r %in% 'FEMALE'))
```

### 3.5.5 Proficiency Level Analysis

The `achievementLevels` function computes the percentages of students' proficiency levels defined by an assessment. The arguments in the function provide user options to

- choose to generate the percentage of individuals performing at each proficiency level (discrete) or at or above each proficiency level (cumulative), and
- calculate the percentage distribution of individuals by proficiency level (discrete or cumulative) and selected characteristics (specified in `aggregateBy`).

```
showCutPoints(ita) # see the proficiency level cut points

# generate the percentage of individuals performing at or above each
# proficiency level (cumulative)
achievementLevels(c('lit'), data = ita, returnDiscrete = FALSE,
returnCumulative = TRUE)

# calculate the percentage distribution of individuals by proficiency level
# (discrete or cumulative) and selected characteristics (specified in
# aggregateBy)
achievementLevels(c('lit', 'gender_r'), data = ita,
aggregateBy = 'gender_r',
returnDiscrete = FALSE,
returnCumulative = TRUE)
```

## 3.6 Notes

1. `getData()`: reads in selected variables and sampling weights from the EdSurvey database and returns a `light.edsurvey.data.frame`. This allows users to work with PIAAC data from other R packages.
2. `addAttributes` is set to the default value of `FALSE`. Setting `addAttributes = TRUE` is one method by which the resultant data object (e.g., `ita`) can be passed to other EdSurvey package functions.
3. All the jackknife replicate weights are returned automatically (`spfw1` to `spfw80`) when a weight is requested.

4. `omittedLevels` is set to TRUE, the default, so that variables with special values, such as multiple entries or NAs, are removed by `getData`. This setting listwise deletes these values from factors that are not typically included in regression analysis and cross-tabulation. Alternatively, `omittedLevels` can be set to FALSE to allow more control by the user.
5. Proficiency levels (PLs) provide an external benchmark against which scale scores can be compared. Six levels—below PL1 to PL5—are defined for each literacy and numeracy, with cut scores for each level determined through a standard-setting process.
6. The `achievementLevels` function applies appropriate weights and the variance estimation method for each `edsurvey.data.frame`, with several arguments for customizing the aggregation and output of the analysis results. By using these optional arguments, users can (a) choose to generate the percentage of individuals performing at each proficiency level (discrete) or at or above each proficiency level (cumulative); (b) calculate the percentage distribution of individuals by proficiency level (discrete or cumulative) and selected characteristics (specified in `aggregateBy`); and (c) compute the percentage distribution of individuals by selected characteristics within a specific proficiency level.

### 3.7 References

- Bailey, P., Lee, M., Nguyen, T., & Zhang, T. (2020). Using EdSurvey to analyse PIAAC data. In D. B. Maehler & B. Rammstedt (Eds.), *Large-scale cognitive assessment: Analyzing PIAAC data*. Springer. [https://doi.org/10.1007/978-3-030-47515-4\\_9](https://doi.org/10.1007/978-3-030-47515-4_9)
- Lee, M., Zhang, T., Bailey, P., Buehler, E., Fink, T., Huo, H., Lee, S., Liao Y., & Sikali, E. (2022). *Analyzing NCES data using EdSurvey: A user's guide*. American Institutes for Research. [https://naep-research.airprojects.org/Portals/0/EdSurvey\\_A\\_Users\\_Guide/\\_book/index.html](https://naep-research.airprojects.org/Portals/0/EdSurvey_A_Users_Guide/_book/index.html)
- Organisation for Economic Cooperation and Development (OECD). (2016). *Programme for the International Assessment of Adult Competencies (PIAAC), Italy Public Use File* [Version: 15433181, prgitap1.sav]. OECD Publishing.

## 4 Analyzing PIAAC Data With Mplus

Author: Ronny Scherer

(Centre for Educational Measurement at the University of Oslo, Norway)

### 4.1 Introduction

PIAAC data files can be analyzed with the statistical software *Mplus* to perform structural equation modeling (SEM), for example. SEM represents a statistical approach to disentangle the relations among latent and/or manifest variables, across groups, over time, and at different analytic levels. This section provides a selection of structural equation models (SEMs) in *Mplus*, including path models with indirect effects, multi-group path models, SEMs with indirect effects, and multi-group SEMs with several invariance constraints (for more information see the [PIAAC scripts digital platform](#)).

### 4.2 Structural Models With PIAAC Data

#### 4.2.1 Path Model CURIOUS-LIFE-PSTRE

The following path model describes the structural relation between participants' curiosity (CURIOUS) and their performance on the PIAAC problem solving in technology-rich environments (PSTRE) items for the Norwegian data from PIAAC Cycle 1 (OECD, 2016). Researchers may hypothesize that this relation is at least partially mediated via the participants' use of skills in everyday life (LIFE). Both curiosity and life skills use are represented as average scale scores, and performance on PSTRE items is represented by the full set of plausible values (PVs).

```
DATA:
FILE IS piaac1-nor_pvlist.dat;
! The file contains the list of names of datasets with one PV each

TYPE = IMPUTATION;
! Needed here as the model contains PVs

VARIABLE:
NAMES ARE
  AGE FEMALE B_Q01a
  D_Q11a D_Q11b D_Q11c D_Q11d
  F_Q03a F_Q03c F_Q05a F_Q05b
  G_Q01a G_Q01b G_Q01c G_Q01d
  G_Q01e G_Q01f G_Q01g G_Q01h
  G_Q02a G_Q02b G_Q02c G_Q02d
  G_Q03b G_Q03c G_Q03d G_Q03f G_Q03g G_Q03h
  G_Q05a G_Q05c G_Q05d G_Q05e
  G_Q05f G_Q05g G_Q05h
  H_Q01a H_Q01b H_Q01c H_Q01d
  H_Q01e H_Q01f H_Q01g H_Q01h
```

```

H_Q02a H_Q02b H_Q02c H_Q02d
H_Q03b H_Q03c H_Q03d H_Q03f H_Q03g H_Q03h
H_Q05a H_Q05c H_Q05d H_Q05e
H_Q05f H_Q05g H_Q05h
I_Q04b I_Q04d I_Q04h I_Q04j I_Q04l I_Q04m
I_Q06a
I_Q07a I_Q07b
HOMLANG IMGEN
PVLIT PVNUM PVPSL
SPFWT0
SPFWT1-SPFWT80
VARSTRAT VARUNIT;

! Variables to be used
! Newly defined variable appear at the end of this list
USEVARIABLES ARE
  PSTRE
  CURIOUS
  LIFE;

! Missing data were coded as -99
MISSING ARE ALL(-99);

! Final participant weight
WEIGHT = SPFWT0;

! Clustering in sampling units
CLUSTER = VARUNIT;

! Stratification
STRATIFICATION = VARSTRAT;

DEFINE:
! Scale down the achievement scores to avoid possible convergence issues
PSTRE = PVPSL/100;
LIT = PVLIT/100;

! Create scale means as composite scores
CURIOUS = (I_Q04b+
I_Q04d+
I_Q04h+
I_Q04j+
I_Q04l+
I_Q04m)/6;

LIFE = (H_Q05a+
H_Q05c+
H_Q05e+
H_Q05f)/4;

```

```

ANALYSIS:
  ! Account for the complex sampling design
  TYPE = COMPLEX;

! Robust maximum-likelihood estimation
  ESTIMATOR = MLR;
  HIITERATIONS = 10000;

MODEL:
  ! STRUCTURAL MODEL
  ! Direct effects with labels a, b, and c
  PSTRE ON
    CURIOUS(c)
    LIFE(b);

  LIFE ON
    CURIOUS(a);

  ! Variance of the exogenous variable
  CURIOUS;

MODEL INDIRECT:
  ! Indirect effect
  PSTRE IND CURIOUS;

MODEL CONSTRAINT:
  ! Estimate indirect and total effects by hand (for comparison)
  new(ind tot);
  ind = a*b;
  tot = ind+c;

OUTPUT:
  STDYX;      ! Fully standardized parameters requested
  STDY;      ! Standardized parameter estimates requested
  SAMPSTAT;  ! Sample statistics
  CINTERVAL; ! Confidence intervals

```

#### 4.2.2 Multi-Group Path Model CURIOUS-LIFE-PSTRE

The code is prepared in the same way as described in [Section 4.2.1](#) (data, variables, definition, and analysis). To extend the path model to a multi-group path model, a grouping-by-gender command is added to the VARIABLE section.

```

VARIABLE:
  [...]
  ! Grouping by gender
  ! Binary coding as 1=Woman and 0=Man
  GROUPING = FEMALE (0=Men 1=Women);

MODEL:
  ! STRUCTURAL MODEL

```

```

! Direct effects
! Note: No more labels here; otherwise, these effects would
! be set equal across groups.
PSTRE ON
    CURIOUS
    LIFE;

LIFE ON
    CURIOUS;

! Variance of the exogenous variable
CURIOUS;

MODEL INDIRECT:
! Indirect effects
PSTRE IND CURIOUS;

OUTPUT:
    STDYX;      ! Fully standardized parameters requested
    STDY;      ! Standardized parameter estimates requested
    SAMPSTAT;  ! Sample statistics
    CINTERVAL; ! Confidence intervals
  
```

### 4.2.3 Multi-Group Path Model CURIOUS-LIFE-PSTRE With Equal $\alpha$ - and $b$ -Paths Across Gender

The code is prepared in the same way as described in [Section 4.2.2](#) (data, variables, definition, and analysis). This model assumes that the  $\alpha$ - and  $b$ -paths in the multi-group path model are the same across gender. Such a model can be compared with a model with the freely estimated paths to test the hypothesis of equal paths.

```

MODEL:
! STRUCTURAL MODEL
! Direct effects
! Labels here to set parameters equal across groups.
PSTRE ON
    CURIOUS(b)
    LIFE;

LIFE ON
    CURIOUS(a);

! Variance of the exogenous variable
CURIOUS;

MODEL INDIRECT:
! Indirect effects
PSTRE IND CURIOUS;
  
```

## OUTPUT:

```

STDYX;           ! Fully standardized parameters requested
STDY;           ! Standardized parameter estimates requested
SAMPSTAT;       ! Sample statistics
CINTERVAL;      ! Confidence intervals

```

#### 4.2.4 Multi-Group Path Model CURIOUS-LIFE-PSTRE With Equal Indirect Effect Across Gender

The code is prepared in the same way as described in [Section 4.2.2](#) (data, variables, definition, and analysis). This model assumes that the indirect effect ( $ab$ ) in the multi-group path model is the same across gender. Such a model can be compared with a model with the freely estimated effect to test the hypothesis of equal effects.

## MODEL:

```

! STRUCTURAL MODEL
! Direct effects
! Note: No more labels in here; otherwise, these effects would
! be set equal across groups.
PSTRE ON
  CURIOUS
  LIFE;

LIFE ON
  CURIOUS;

! Variance of the exogenous variable
CURIOUS;

MODEL Men:
! Model with group-specific labels of paths
PSTRE ON
  CURIOUS
  LIFE(bM);

LIFE ON
  CURIOUS(aM);

! Variance of the exogenous variable
CURIOUS;

MODEL Women:
! Model with group-specific labels of paths
PSTRE ON
  CURIOUS
  LIFE(bW);

LIFE ON
  CURIOUS(aW);

```

```

! Variance of the exogenous variable
CURIOUS;

MODEL INDIRECT:
! Indirect effects
PSTRE IND CURIOUS;

MODEL CONSTRAINT:
! Set the indirect effects equal across gender
0 = aM*bM - aW*bW;

OUTPUT:
          STDYX;          ! Fully standardized parameters requested
          STDY;          ! Standardized parameter estimates requested
          SAMPSTAT;      ! Sample statistics
          CINTERVAL;     ! Confidence intervals

```

### 4.3 Structural Equation Models With PIAAC Data

Except for the inclusion of latent variables, the structural part of these SEMs is identical with that of the path model with manifest variables. As a result, the model syntax is modified by adding the measurement models of curiosity and skills use.

#### 4.3.1 SEM With Measurement Models and Complex Structural Relations

This SEM assumes several structural relations among curiosity, life skills use in everyday life, performance on PSTRE items, and participants' background. Besides the measurement models of life skills and curiosity, the structural model contains the relations between the manifest variables, the latent variables, and the manifest and latent variables. Notably, this example code contains an effect (gender) that describes possible (uniform) differential item functioning (DIF).

```

DATA:
FILE IS piaac1-nor_pvlist.dat;
! The file contains the list of names of datasets with one PV each

TYPE = IMPUTATION;
! Needed here as the model contains PVs

VARIABLE:
NAMES ARE
AGE FEMALE B_Q01a
D_Q11a D_Q11b D_Q11c D_Q11d
F_Q03a F_Q03c F_Q05a F_Q05b
G_Q01a G_Q01b G_Q01c G_Q01d
G_Q01e G_Q01f G_Q01g G_Q01h
G_Q02a G_Q02b G_Q02c G_Q02d
G_Q03b G_Q03c G_Q03d G_Q03f G_Q03g G_Q03h
G_Q05a G_Q05c G_Q05d G_Q05e

```



```

G_Q05f G_Q05g G_Q05h
H_Q01a H_Q01b H_Q01c H_Q01d
H_Q01e H_Q01f H_Q01g H_Q01h
H_Q02a H_Q02b H_Q02c H_Q02d
H_Q03b H_Q03c H_Q03d H_Q03f H_Q03g H_Q03h
H_Q05a H_Q05c H_Q05d H_Q05e
H_Q05f H_Q05g H_Q05h
I_Q04b I_Q04d I_Q04h I_Q04j I_Q04l I_Q04m
I_Q06a
I_Q07a I_Q07b
HOMLANG IMGEN
PVLIT PVNUM PVPSL
SPFWT0
SPFWT1-SPFWT80
VARSTRAT VARUNIT;

! Variables to be used
! Newly defined variable appear at the end of this list
USEVARIABLES ARE
  H_Q05a
  H_Q05c
  H_Q05e
  H_Q05f
  I_Q04b
  I_Q04d
  I_Q04h
  I_Q04j
  I_Q04l
  I_Q04m
  HOMLANG
  FEMALE
  LIT
  PSTRE;

! Missing data are coded as -99
MISSING ARE ALL(-99);

! Final participant weight
WEIGHT = SPFWT0;

! Clustering in sampling units
CLUSTER = VARUNIT;

! Stratification
STRATIFICATION = VARSTRAT;

DEFINE:
! Scale down the achievement scores to avoid possible convergence issues
PSTRE = PVPSL/100;
LIT = PVLIT/100;

```

```

ANALYSIS:
  ! Account for the complex sampling design
  TYPE = COMPLEX;

! Robust maximum-likelihood estimation
ESTIMATOR = MLR;
HIITERATIONS = 10000;

MODEL:
  ! MEASUREMENT MODELS
  ! Life skills use in every-day life
  LIFE BY
    H_Q05a
    H_Q05c
    H_Q05e
    H_Q05f;

  ! Curiosity
  CURIOUS BY
    I_Q04b
    I_Q04d
    I_Q04h
    I_Q04j
    I_Q04l
    I_Q04m;

  ! Covariances between residuals (beyond the latent variable)
  I_Q04B WITH I_Q04H;
  I_Q04B WITH I_Q04J;
  H_Q05E WITH H_Q05F;

  ! STRUCTURAL MODEL
  ! Direct effects
  PSTRE ON
    CURIOUS
    LIFE
    FEMALE
    HOMLANG
    LIT;

  CURIOUS ON
    FEMALE
    HOMLANG
    LIT;

  LIFE ON
    FEMALE
    HOMLANG
    LIT;

  ! Variances of the exogenous variables and mediator residuals
  CURIOUS;

```

```

LIFE;
FEMALE;
HOMLANG;
LIT;

! Covariances among exogenous variables and mediator residuals
! Residuals of mediators
CURIOUS WITH LIFE;
! Predictors
FEMALE WITH HOMLANG LIT;
HOMLANG WITH LIT;

! Uniform DIF effect
H_Q05E ON FEMALE;

MODEL INDIRECT:
! Indirect effects
PSTRE IND FEMALE;
PSTRE IND HOMLANG;
PSTRE IND LIT;

OUTPUT:
          STDYX;           ! Fully standardized parameters requested
          STDY;            ! Standardized parameter estimates requested
          SAMPSTAT;       ! Sample statistics
          CINTERVAL;     ! Confidence intervals

```

### 4.3.2 Multi-Group SEM With Measurement Models, Complex Structural Relations, and Configural Measurement and Regression Invariance

Although the basic setup of this code example follows the example in [Section 4.3.1](#) (data, variables, model), it has some extensions. Specifically, gender (FEMALE) is introduced as a grouping variable, and no longer appears as an explicit predictor variable in the MODEL section. In this multi-group SEM, we assume *configural measurement and regression invariance*—that is, all parameters in the measurement and structural model are freely estimated across gender, and the same number of latent variables and loading patterns is assumed.

```

VARIABLE:
[...]

! Grouping by gender
GROUPING IS FEMALE (1=Women 0=Men);

DEFINE:
! Scale down the achievement scores to avoid possible convergence issues
PSTRE = PVPSL/100;
LIT = PVLIT/100;

```

```

ANALYSIS:
  ! Account for the complex sampling design
  TYPE = COMPLEX;

  ! Robust maximum-likelihood estimation
  ESTIMATOR = MLR;
  HIITERATIONS = 10000;

MODEL:
  ! MEASUREMENT MODELS
  ! Everyday life skills
  LIFE BY
    H_Q05a
    H_Q05c
    H_Q05e
    H_Q05f;

  ! Curiosity
  CURIOUS BY
    I_Q04b
    I_Q04d
    I_Q04h
    I_Q04j
    I_Q04l
    I_Q04m;

  ! Covariances among residuals
  I_Q04B WITH I_Q04H;
  I_Q04B WITH I_Q04J;
  H_Q05E WITH H_Q05F;

  ! STRUCTURAL MODEL
  ! Direct effects
  PSTRE ON
    CURIOUS
    LIFE
    HOMLANG
    LIT;

  CURIOUS ON
    HOMLANG
    LIT;

  LIFE ON
    HOMLANG
    LIT;

  ! Variances of predictors and residuals of mediators
  CURIOUS;
  LIFE;
  HOMLANG;
  LIT;

```

```

! Covariances among predictors and residuals of mediators
! Residuals of mediators
CURIOUS WITH LIFE;
! Predictors
HOMLANG WITH LIT;

MODEL WOMEN:
! MEASUREMENT MODELS gender-specific
! Everyday life skills
LIFE BY
    H_Q05a
    H_Q05c
    H_Q05e
    H_Q05f;

! Curiosity
CURIOUS BY
    I_Q04b
    I_Q04d
    I_Q04h
    I_Q04j
    I_Q04l
    I_Q04m;

! Covariances among residuals
I_Q04B WITH I_Q04H;
I_Q04B WITH I_Q04J;
H_Q05E WITH H_Q05F;

! STRUCTURAL MODEL gender-specific
! Direct effects
PSTRE ON
    CURIOUS
    LIFE
    HOMLANG
    LIT;

CURIOUS ON
    HOMLANG
    LIT;

LIFE ON
    HOMLANG
    LIT;

! Variances of predictors and residuals of mediators
CURIOUS;
LIFE;
HOMLANG;
LIT;

! Covariances among predictors and residuals of mediators

```

```
! Residuals of mediators
```

```
CURIOUS WITH LIFE;
```

```
! Predictors
```

```
HOMLANG WITH LIT;
```

```
MODEL MEN:
```

```
! MEASUREMENT MODELS gender-specific
```

```
! Everyday life skills
```

```
LIFE BY
```

```
    H_Q05a
```

```
    H_Q05c
```

```
    H_Q05e
```

```
    H_Q05f;
```

```
! Curiosity
```

```
CURIOUS BY
```

```
    I_Q04b
```

```
    I_Q04d
```

```
    I_Q04h
```

```
    I_Q04j
```

```
    I_Q04l
```

```
    I_Q04m;
```

```
! Covariances among residuals
```

```
I_Q04B WITH I_Q04H;
```

```
I_Q04B WITH I_Q04J;
```

```
H_Q05E WITH H_Q05F;
```

```
! STRUCTURAL MODEL gender-specific
```

```
! Direct effects
```

```
PSTRE ON
```

```
    CURIOUS
```

```
    LIFE
```

```
    HOMLANG
```

```
    LIT;
```

```
CURIOUS ON
```

```
    HOMLANG
```

```
    LIT;
```

```
LIFE ON
```

```
    HOMLANG
```

```
    LIT;
```

```
! Variances of predictors and residuals of mediators
```

```
CURIOUS;
```

```
LIFE;
```

```
HOMLANG;
```

```
LIT;
```

```
! Covariances among predictors and residuals of mediators
```

```
! Residuals of mediators
```

```
CURIOUS WITH LIFE;
```

```
! Predictors
```

```
HOMLANG WITH LIT;
```

```
OUTPUT:
```

```
STDYX;           ! Fully standardized parameters requested
STDY;           ! Standardized parameter estimates requested
SAMPSTAT;       ! Sample statistics
CINTERVAL;      ! Confidence intervals
```

### 4.3.3 Multi-Group SEM With Measurement Models, Complex Structural Relations, and Metric Measurement and Regression Invariance

The data preparation (data, variables, definition and analysis) is the same as that described in [Section 4.3.2](#). This example adds the assumption of metric invariance to the measurement models (i.e., equal factor loadings across gender).

```
MODEL:
```

```
! MEASUREMENT MODELS
```

```
! Everyday life skills
```

```
! Equality constraints to the model with labels L2-L4
```

```
LIFE BY
```

```
  H_Q05a
```

```
  H_Q05c(L2)
```

```
  H_Q05e(L3)
```

```
  H_Q05f(L4);
```

```
! Curiosity
```

```
! Equality constraints to the model with labels L5-L9
```

```
CURIOUS BY
```

```
  I_Q04b
```

```
  I_Q04d(L5)
```

```
  I_Q04h(L6)
```

```
  I_Q04j(L7)
```

```
  I_Q04l(L8)
```

```
  I_Q04m(L9);
```

```
! Covariances among residuals
```

```
I_Q04B WITH I_Q04H;
```

```
I_Q04B WITH I_Q04J;
```

```
H_Q05E WITH H_Q05F;
```

```
! STRUCTURAL MODEL
```

```
! Direct effects
```

```
PSTRE ON
```

```
  CURIOUS
```

```
  LIFE
```

```
  HOMLANG
```

```
  LIT;
```

```

CURIOUS ON
    HOMLANG
    LIT;

LIFE ON
    HOMLANG
    LIT;

! Variances of predictors and residuals of mediators
CURIOUS;
LIFE;
HOMLANG;
LIT;

! Covariances among predictors and residuals of mediators
! Residuals of mediators
CURIOUS WITH LIFE;
! Predictors
HOMLANG WITH LIT;

MODEL WOMEN:
! Covariances among residuals
I_Q04B WITH I_Q04H;
I_Q04B WITH I_Q04J;
H_Q05E WITH H_Q05F;

! STRUCTURAL MODEL gender-specific
! Direct effects
PSTRE ON
    CURIOUS
    LIFE
    HOMLANG
    LIT;

CURIOUS ON
    HOMLANG
    LIT;

LIFE ON
    HOMLANG
    LIT;

! Variances of predictors and residuals of mediators
CURIOUS;
LIFE;
HOMLANG;
LIT;

! Covariances among predictors and residuals of mediators
! Residuals of mediators
CURIOUS WITH LIFE;
! Predictors

```



```

    HOMLANG WITH LIT;

MODEL MEN:
  ! Covariances among residuals
  I_Q04B WITH I_Q04H;
  I_Q04B WITH I_Q04J;
  H_Q05E WITH H_Q05F;

  ! STRUCTURAL MODEL gender-specific
  ! Direct effects
  PSTRE ON
    CURIOUS
    LIFE
    HOMLANG
    LIT;

  CURIOUS ON
    HOMLANG
    LIT;

  LIFE ON
    HOMLANG
    LIT;

  ! Variances of predictors and residuals of mediators
  CURIOUS;
  LIFE;
  HOMLANG;
  LIT;

  ! Covariances among predictors and residuals of mediators
  ! Residuals of mediators
  CURIOUS WITH LIFE;
  ! Predictors
  HOMLANG WITH LIT;

OUTPUT:
  STDYX;      ! Fully standardized parameters requested
  STDY;       ! Standardized parameter estimates requested
  SAMPSTAT;   ! Sample statistics
  CINTERVAL;  ! Confidence intervals

```

#### 4.3.4 Multi-Group SEM With Measurement Models, Complex Structural Relations, and Metric Measurement and Configural Regression Invariance

The data preparation (data, variables, definition and analysis) is the same as that described in [Section 4.3.3](#). This example adds the assumption of regression invariance to the structural models (i.e., equal path coefficients across gender).

```

MODEL:
  ! MEASUREMENT MODELS
  ! Everyday life skills
  ! Equality constraints to the model with labels L2-L4
LIFE BY
  H_Q05a
  H_Q05c(L2)
  H_Q05e(L3)
  H_Q05f(L4);

  ! Curiosity
  ! Equality constraints to the model with labels L5-L9
CURIOUS BY
  I_Q04b
  I_Q04d(L5)
  I_Q04h(L6)
  I_Q04j(L7)
  I_Q04l(L8)
  I_Q04m(L9);

  ! Covariances among residuals
I_Q04B WITH I_Q04H;
I_Q04B WITH I_Q04J;
H_Q05E WITH H_Q05F;

  ! STRUCTURAL MODEL
  ! Direct effects
PSTRE ON
  CURIOUS(R1)
  LIFE(R2)
  HOMLANG(R3)
  LIT(R4);

CURIOUS ON
  HOMLANG(R5)
  LIT(R6);

LIFE ON
  HOMLANG(R7)
  LIT(R8);

  ! Variances of predictors and residuals of mediators
CURIOUS;
LIFE;
HOMLANG;
LIT;

  ! Covariances among predictors and residuals of mediators
  ! Residuals of mediators
CURIOUS WITH LIFE;
  ! Predictors
HOMLANG WITH LIT;

```

```
MODEL WOMEN:
  ! Covariances among residuals
  I_Q04B WITH I_Q04H;
  I_Q04B WITH I_Q04J;
  H_Q05E WITH H_Q05F;

  ! Variances of predictors and residuals of mediators
  CURIOUS;
  LIFE;
  HOMLANG;
  LIT;

  ! Covariances among predictors and residuals of mediators
  ! Residuals of mediators
  CURIOUS WITH LIFE;
  ! Predictors
  HOMLANG WITH LIT;

MODEL MEN:
  ! Covariances among residuals
  I_Q04B WITH I_Q04H;
  I_Q04B WITH I_Q04J;
  H_Q05E WITH H_Q05F;

  ! Variances of predictors and residuals of mediators
  CURIOUS;
  LIFE;
  HOMLANG;
  LIT;

  ! Covariances among predictors and residuals of mediators
  ! Residuals of mediators
  CURIOUS WITH LIFE;
  ! Predictors
  HOMLANG WITH LIT;

OUTPUT:
  STDYX;           ! Fully standardized parameters requested
  STDY;           ! Standardized parameter estimates requested
  SAMPSTAT;       ! Sample statistics
  CINTERVAL;      ! Confidence intervals
```

## 4.4 References

- Scherer, R. (2020). Analysing PIAAC data with structural equation modelling in Mplus. In D. B. Maehler & B. Rammstedt (Eds.), *Large-scale cognitive assessment* (pp. 165–208). Springer. [https://doi.org/10.1007/978-3-030-47515-4\\_8](https://doi.org/10.1007/978-3-030-47515-4_8)
- Organisation for Economic Cooperation and Development (OECD). (2016). *Programme for the International Assessment of Adult Competencies (PIAAC), Norway Public Use File* [Version: 17723269, prgnorp1.sav]. OECD Publishing.

## 5 Using the IDB Analyzer to Analyze PIAAC Data With SPSS, SAS, and R

*Authors: Umut Atasever & Rolf Strietholt  
(IEA – International Association for the Evaluation of Educational Achievement, Germany)*

### 5.1 Introduction

Because PIAAC uses a complex assessment and a clustered sampling design, PIAAC data cannot be analyzed correctly with SPSS, SAS, or R without using specialized packages or macros. Sample statistics and the corresponding standard errors cannot be calculated in a conventional way. Rather, various features of the complex study design must be taken into account (see Gonzalez, 2014, for more information on sampling weights, replication weights, and plausible values). In addition, as sampling designs in PIAAC vary across countries, different methods must be used in each country to estimate sampling variance (see Mohadjer et al., 2013).

The IEA International Database Analyzer (IDB Analyzer) is a software tool developed by the International Association for the Evaluation of Educational Achievement (IEA). It can be used to analyze data from PIAAC and other international large-scale assessments while accounting for sampling and replication weights as well as plausible values. The user-friendly graphical interface generates syntax code for SPSS, SAS, or R to combine data files and conduct analyses with these data. It can be used to compute percentages, means, correlations, and regression analyses while fully taking into account the complex PIAAC design. The software identifies what data are being used and selects appropriate methods and procedures for the data analyses. All procedures correctly use the sampling weights and compute standard errors according to the variance estimation of the corresponding study and country. Analyses that include plausible values for the performance tests—either as dependent or independent variables—consider all plausible values when calculating standard errors. In the following sections, we illustrate the code generated by the IDB Analyzer using mean and regression analyses as examples, and we show code for SPSS, SAS, and R (for detailed information on how the IDB Analyzer works, see the [PIAAC scripts digital platform](#)).

It should be noted that the analyses themselves are not performed using the IDB Analyzer but rather within the respective software package. The analyses must therefore be started within that software package (e.g., by using Ctrl+R in SPSS). After running the analyses, the results are presented in the file formats of the corresponding analysis software, as well as in generic formats such as HTML, CSV or MS Excel XLSX files. The IDB Analyzer can be downloaded free of charge from the IEA website ([www.iea.nl](http://www.iea.nl)), which also offers video tutorials and a regularly updated help manual (IEA, 2022).

### 5.2 Computing Means With Plausible Values

The following syntaxes were created by the IDB Analyzer to compute means of literacy performance grouped by the grouping variables country (CNTRYID) and migration background (IMGEN). As there are 10 plausible values for performance in literacy (PVLIT1 to PVLIT10), the

analyses will be replicated 10 times and then combined (for detailed information see the [PIAAC scripts digital platform](#)).

### 5.2.1 SPSS

The following code for SPSS was generated using the IDB Analyzer to estimate mean values per group.

```
* Script created using the IEA IDB Analyzer (Version 5.0.16).
* Created on 1/15/2023 at 6:57 PM.
* Press Ctrl+A followed by Ctrl+R to submit this analysis.

include file = "C:\Users\umut.atasever\AppData\Roaming\IEA\IDBAnalyzerV5\bin\
Data\Templates\SPSS_Macros\JB_PV.ieasps".

JB_PV      infile="C:\PIAAC_Analysis\00_Data\PIAAC_PRG_merged.sav"/
          cvar=CNTRYID IMGEN/
          almvars=/
          rootpv=PVLIT /
          tailpv=/
          npv=10/
          wgt=SPFWT0/
          nrwt=80 /
          rwgt=SPFWT/
          jkz=/
          jkr=/
          jk2type=HALF/
          stratvar=/
          nomiss=Y/
          method=PIAAC/
          kfac=0/
          shrtcut=N/
          viewcod=N/
          ndec=2/
          clean = Y/
          strctry = N/
          intavg = Y/
          graphs=Y/
          selcrit = /
          selvar = /
          outdir="C:\PIAAC_Analysis\01_Syntax"/
          outfile="Means_IMGEN".
```

### 5.2.2 SAS

The following code for SAS was generated using the IDB Analyzer to estimate mean values per group.

```
/* Script created using the IEA IDB Analyzer (Version 5.0.16). */
/* Created on 1/15/2023 at 7:05 PM. */
```

```

/* Press F8 to submit this analysis. */

/* This is where the macros are located */
%let mdir = C:\Users\umut.atasever\AppData\Roaming\IEA\IDBAnalyzerV5\bin\Data\
Templates\SAS_Macros;

/* This is where the data will be read from */
%let idir = C:\PIAAC_Analysis\00_Data;

/* This is where the output will be saved */
%let odir = C:\PIAAC_Analysis\01_Syntax;

/* DO NOT EDIT AFTER THIS */
options mrecall sasautos = (sasautos "&mdir");

%jb_pv ( InDir = &idir ,
        InFile = PIAAC_PRG_merged ,
        SelVar = ,
        SelCrit = ,
        OutDir = &odir ,
        OutFile = Means_IMGEN ,
        CVar = CNTRYID IMGEN ,
        Almvars = ,
        NoMiss = Y ,
        strctry = N ,
        rootpv = PVLIT ,
        tailpv = ,
        shrtcut = N ,
        npv = 10 ,
        method = PIAAC ,
        kfac = 0 ,
        wgt = SPFWTO ,
        rwgt = SPFWT ,
        nrwgt = 80 ,
        jkz = ,
        jkr = ,
        jk2type = HALF ,
        stratvar = ,
        newout = Y ,
        qcstats = Y ,
        report = Y ,
        graphs = Y ,
        intavg = Y ,
        viewlbl = Y ,
        ndec = 2 ,
        viewcod = N ,
        ViewPrgs = Y ,
        clean = Y );

```

### 5.2.3 R

The following code for R was generated using the IDB Analyzer to estimate mean values per group.

```
# Percentages and Means: JB_PV.R

# =====
# Script created using the IEA IDB Analyzer (Version 5.0.16).
# Created on 1/15/2023 at 6:13 PM.
# Press Ctrl+A followed by Ctrl+Enter to submit this analysis.

#####

# IEA IDB Analyzer: R Module
# Programmer: IEA Hamburg, please contact idb-analyzer@iea-hamburg.de
#####

# Overall function
# Clean workspace and define settings =====
=====

rm(list = ls())
# Directory, where the IEA IDB Analyzer macros are stored
include_file <- "C:/Users/umut.atasever/AppData/Roaming/IEA/IDBAnalyzerV5/bin/
Data/Templates/R_Macros"

source(sprintf("%s/check_packages.R", include_file), local = TRUE)
source(sprintf("%s/JB_PV.R", include_file), local = TRUE)

# Dependencies =====
==

library(dplyr)      # version 1.0.10
library(ggplot2)   # version 3.3.6
library(haven)     # version 2.5.1
library(htmltools) # version 0.5.3
library(kableExtra) # version 1.3.4
library(knitr)     # version 1.40
library(openxlsx)  # version 4.2.5.1
library(rmarkdown) # version 2.17
library(sjlabelled) # version 1.2.0
library(tidyr)     # version 1.2.1
library(tidyselect) # version 1.2.0

# =====

JB_PV(
  infile = "C:/PIAAC_Analysis/00_Data/PIAAC_PRG_MERGED.Rdata",
  cvar = c("CNTRYID", "IMGEN"),
  almvars = NULL,
  rootpv = c("PVLIT"),
  tailpv = NULL,
```



```

npv = 10,
wgt = „SPFWTO“,
nrwgt = 80,
rwgt = „SPFWT“,
jkz = NULL,
jkr = NULL,
jk2type = “HALF”,
stratvar = NULL,
nomiss = “Y”,
method = “PIAAC”,
kfac = 0,
shrtcut = “N”,
viewcod = “N”,
ndec = 2,
clean = “Y”,
intavg = “Y”,
graphs = “Y”,
qcstats = “N”,
report = “Y”,
selcrit = NULL,
selvar = NULL,
outdir = “C:/PIAAC_Analysis/01_Syntax”,
outfile = “Means_IMPAR”)

```

### 5.3 Computing Linear Regression With Plausible Values As An Explanatory Variable

The following syntaxes were created by the IDB Analyzer to conduct ordinary least squares (OLS) regression with plausible values as an independent variable. The dependent variable—use of reading skills at work (READWORK)—was regressed on the independent variables immigration background (IMGEN), social trust (I\_Q07a), and literacy performance. As there are 10 plausible values for performance in literacy (PVLIT1 to PVLIT10), the analyses will be replicated 10 times and then combined. The immigration background variable is dummy coded using the first category as the reference category; the social trust variable is effect coded using the third category as the reference category. Regression models were estimated separately for the different countries (for detailed information see the [PIAAC scripts digital platform](#)).

#### 5.3.1 SPSS

The following code for SPSS was generated using the IDB Analyzer to estimate an OLS regression model.

```

* Script created using the IEA IDB Analyzer (Version 5.0.16).
* Created on 1/24/2023 at 5:50 PM.
* Press Ctrl+A followed by Ctrl+R to submit this analysis.

```

```

include file = “C:\Users\umut.atasever\AppData\Roaming\IEA\IDBAnalyzerV5\bin\
Data\Templates\SPSS_Macros\JB_RegGP.ieasps”.

```

```

JB_RegGP      infile="C:\PIAAC_Analysis\00_Data\PIAAC_PRG_merged.sav"/
              cvar=CNTYID /
              convar=/
              catvar=IMGEN I_Q07A /
              codings=D E/
              refcats=1 3/
              ncats=3 5/
              PVRoots=PVLIT /
              PVTails=/
              dvar0=READWORK /
              rootpv=/
              tailpv=/
              npv=10/
              wgt=SPFWT0/
              nrwt=80 /
              rwt=SPFWT/
              jkz=/
              jkr=/
              jk2type=HALF/
              stratvar=/
              nomiss=Y/
              method=PIAAC/
              missing=listwise/
              kfac=0/
              shrtcut=N/
              viewcod=N/
              ndec=2/
              clean = Y/
              strctry = N/
              viewprgs=Y/
              viewlbl=Y/
              qcstats=Y/
              newout=Y/
              intavg = Y/
              selcrit = /
              selvar = /
              outdir="C:\PIAAC_Analysis\01_Syntax"/
              outfile="Regression_IMGEN_TRUST_LITERACY_READWORK".
  
```

### 5.3.2 SAS

The following code for SAS was generated using the IDB Analyzer to estimate an OLS regression model.

```

/* Script created using the IEA IDB Analyzer (Version 5.0.16). */
/* Created on 1/24/2023 at 5:56 PM. */
/* Press F8 to submit this analysis. */

/* This is where the macros are located */
%let mdir = C:\Users\umut.atasever\AppData\Roaming\IEA\IDBAnalyzerV5\bin\Data\
Templates\SAS_Macros;
  
```

```

/* This is where the data will be read from */
%let idir = C:\PIAAC_Analysis\00_Data;

/* This is where the output will be saved */
%let odir = C:\PIAAC_Analysis\01_Syntax;

/* DO NOT EDIT AFTER THIS */
options mrecall sasautos = (sasautos "&mdir");

%jb_reggp (
  InDir = &idir ,
  InFile = PIAAC_PRG_merged ,
  SelVar = ,
  SelCrit = ,
  OutDir = &odir ,
  OutFile = Regression_IMGEN_TRUST_LITERACY_READWORK ,
  CVar = CNTRYID ,
  NoMiss = Y ,
  strctry = N ,
  dvar0 = READWORK ,
  catvar = IMGEN I_Q07A ,
  codings = D E ,
  ncats = 3 5 ,
  RefCats = 1 3 ,
  convar = ,
  PVRoots = PVLIT ,
  PVTails = ,
  rootpv = ,
  tailpv = ,
  shrtcut = N ,
  npv = 10 ,
  method = PIAAC ,
  kfacs = 0 ,
  wgt = SPFWT0 ,
  rwgt = SPFWT ,
  nrwgt = 80 ,
  jkz = ,
  jkr = ,
  jk2type = HALF ,
  stratvar = ,
  missing = listwise ,
  newout = Y ,
  qcstats = Y ,
  Report = Y ,
  intavg = Y ,
  viewlbl = Y ,
  ndec = 2 ,
  viewcod = N ,
  ViewPrs = Y ,
  clean = Y );

```

### 5.3.3 R

The following code for R was generated using the IDB Analyzer to estimate an OLS regression model.

```
# Linear Regression: JB_RegGP

# =====

# Script created using the IEA IDB Analyzer (Version 5.0.16).
# Created on 1/24/2023 at 5:58 PM.
# Press Ctrl+A followed by Ctrl+Enter to submit this analysis.

#####
#
# IEA IDB Analyzer: R Module
# Programmer: IEA Hamburg, please contact idb-analyzer@iea-hamburg.de
#
#####

# Overall function

# Clean workspace and define settings =====
=====

rm(list = ls())

# Directory, where the IEA IDB Analyzer macros are stored
include_file <- "C:/Users/umut.atasever/AppData/Roaming/IEA/IDBAnalyzerV5/bin/
Data/Templates/R_Macros"

source(sprintf("%s/check_packages.R", include_file), local = TRUE)
source(sprintf("%s/JB_RegGP.R", include_file), local = TRUE)

# Dependencies =====
=====

library(dplyr)      # version 1.0.10
library(ggplot2)   # version 3.3.6
library(haven)     # version 2.5.1
library(htmltools) # version 0.5.3
library(kableExtra) # version 1.3.4
library(knitr)     # version 1.40
library(openxlsx)  # version 4.2.5.1
library(rmarkdown) # version 2.17
library(sjlabelled) # version 1.2.0
library(tidyr)     # version 1.2.1
library(tidyselect) # version 1.2.0

# =====
```

```

JB_RegGP(  infile = "C:/PIAAC_Analysis/00_Data/PIAAC_PRG_MERGED.Rdata",
          cvar = c("CNTRYID"),
          convar = NULL,
          catvar = c("IMGEN","I_Q07A"),
          codings = c(„D“,“E“),
          refcats = c(1,3),
          ncats = c(3,5),
          PVRoots = c("PVLIT"),
          PVTails = NULL,
          dvar0 = "READWORK",
          rootpv = NULL,
          tailpv = NULL,
          npv = 10,
          wgt = „SPFWTO“,
          nrwgt = 80,
          rwgt = „SPFWT“,
          jkz = NULL,
          jkr = NULL,
          jk2type = "HALF",
          stratvar = NULL,
          nomiss = "Y",
          method = "PIAAC",
          missing = "listwise",
          kfac = 0,
          shrtcut = "N",
          viewcod = "N",
          ndec = 2,
          clean = "Y",
          qcstats = "Y",
          intavg = "Y",
          selcrit = NULL,
          selvar = NULL,
          outdir = "C:/PIAAC_Analysis/01_Syntax",
          outfile = "Regression_IMGEN_TRUST_LITERACY_READWORK")

```

## 5.4 References

- Gonzales, E. J. (2014). Calculating standard errors of sample statistics when using international large-scale assessment data. In R. Strietholt, W. Bos, J. E. Gustafsson, & M. Rosén (Eds.), *Educational policy evaluation through international comparative assessments* (pp. 59–74). Waxmann.
- International Association for the Evaluation of Educational Achievement (IEA). (2022). *Help manual for the IEA IDB Analyzer* (Version 5.0). IEA. <https://www.iea.nl/sites/default/files/2022-06/IDB-Analyzer-Manual-%28Version-5-0%29.pdf>
- Mohadjer, L., Krenzke, T., Van de Kerckhove, W., & Hsu, V. (2013). Survey weighting and variance estimation. In OECD (Ed.), *Technical report of the Survey of Adult Skills (PIAAC)* (Chapter 15). Organisation for Economic Co-operation and Development (OECD). [https://www.oecd.org/skills/piaac/\\_Technical%20Report\\_17OCT13.pdf](https://www.oecd.org/skills/piaac/_Technical%20Report_17OCT13.pdf)



if (Large\_Scale\_Assesment(2022))

<data>

<M-plus>

<2012>=<Research>

var\_13 <PIAAC>

var\_C+

va4\_4<data>

<PIAAC//Research>

<R>

<Stata>

att\_8\_vR base frame 1x

<SPSS>  
string\_to\_string

<\_Information>\_<Analyses>

<\_International\_PIAAC\_Research >

if (International\_Research(2023))

<Large-scale Assesment.2024>

SPONSORED BY THE



Federal Ministry  
of Education  
and Research

gesis

Leibniz Institute  
for the Social Sciences